From Heritage to Horizon: Language, Literature, and Identity in the Digital Preservation of Regional Cultural Narratives

Dr. Vaishaliben Randhirsinh Borasiya
Assistant professor (English)
SDJ International College Palsana, Gujarat.
Email:- vansia.vaishali31@gmail.com

Abstract:

The rapid erosion of linguistic and cultural diversity in the digital age presents an urgent challenge to the sustainability of regional identities. This paper examines the intersection of language, literature, and identity through the lens of digital preservation of regional cultural narratives. Drawing upon a qualitative, multi-case study methodology, the research investigates how digital technologies—ranging from community-driven archives to AI-powered language revitalisation tools—are being deployed to sustain endangered linguistic and literary traditions. The study analyses diverse initiatives including the REVIVE project for Cornish and Griko languages, the Pangloss Collection, the BHASHINI platform in India, and community-centred digital heritage studios. Key findings reveal that technology functions as an enabler rather than a solution in itself, with sustainable preservation contingent upon active community participation, ethical data governance, and the integration of oral traditions into digital ecosystems. The paper proposes a conceptual framework for "heritage-to-horizon" sustainability, positioning digital preservation not merely as archival practice but as a dynamic process of cultural negotiation and identity re-articulation. The findings contribute to the United Nations Sustainable Development Goals, particularly SDG 4 (Quality Education), SDG 9 (Industry, Innovation and Infrastructure), and SDG 11 (Sustainable Cities and Communities).

1. Introduction:

Language is not merely a means of communication; it is the foundational architecture of cultural identity, a living repository of indigenous knowledge, and the medium through which literature and collective memory are transmitted across generations. The world is currently witnessing an unprecedented erosion of linguistic diversity, with UNESCO estimating that nearly half of the world's approximately 7,000 languages face the threat of extinction (UNESCO, 2026). This crisis is not solely a matter of linguistic loss; it represents an irreversible diminution of the cognitive and cultural diversity that has sustained human societies for millennia.

The digital age presents a paradox. On one hand, globalisation and digital homogenisation have accelerated the marginalisation of minority languages, with digital platforms overwhelmingly privileging dominant linguistic and cultural frameworks. On the other hand, the same digital technologies—when deployed ethically and collaboratively—offer transformative potential for documenting, revitalising, and sustaining endangered linguistic and literary heritage. As noted by the Collaborative Digital Heritage Studio at the University of British Columbia, "languages are at risk, traditions fade, and cultural knowledge becomes

endangered, sometimes faster than we realize", yet communities worldwide are harnessing new technologies to resist these changes in ethical, collaborative ways.

This paper addresses the research problem of how digital preservation initiatives can effectively contribute to the sustainability of regional cultural narratives—understood as the interwoven fabric of language, literature, and identity. The central research question guiding this investigation is: How can digital technologies be deployed to preserve regional languages and literary traditions in ways that are both culturally sustainable and identity-affirming?

The study operates within the theoretical parameters of "heritage-to-horizon" thinking, a framework that positions cultural preservation not as a backward-looking exercise in nostalgia but as a forward-oriented process of innovation, adaptation, and sustainable cultural transmission. The conference theme of "Innovation for a Sustainable Future" finds direct expression in this approach, which seeks to bridge the past and the future through technologically mediated cultural practice.

The paper proceeds as follows: Section 2 provides a review of relevant literature on digital preservation, language endangerment, and cultural identity. Section 3 outlines the methodology. Section 4 presents key case studies. Section 5 discusses findings and proposes a conceptual framework. Section 6 concludes with implications for policy and practice.

2. Literature Review

2.1 Language Endangerment and the Digital Divide

The phenomenon of language endangerment has been extensively documented across sociolinguistics and anthropological linguistics. Languages become endangered through a complex interplay of factors including colonisation, urbanisation, intergenerational disruption, and the economic and social pressures of linguistic homogenisation. Digital technology has historically exacerbated this process, creating what scholars term a "digital divide" that disproportionately affects speakers of minority languages. As one study of the Irula tribe in India's Western Ghats observes, the digital divide's impact on preserving endangered languages is profound, with low digital literacy and inadequate technology access representing significant barriers.

However, this narrative of digital technology as exclusively destructive is incomplete. The same study demonstrates that digital storytelling, integrated into a Mobile-Assisted Language Learning (MALL) framework, can support language revitalisation when grounded in situated learning and social constructivist theories. The key insight is that technology must

be culturally tailored, with indigenous knowledge—oral history and folklore—embedded into digital platforms through community-driven processes.

2.2 Digital Preservation: From Archiving to Revitalisation

Traditional approaches to digital preservation have focused predominantly on documentation—the recording of grammatical structures, vocabulary, and lexicons. As the REVIVE project observes, "previous efforts have predominantly concentrated on preserving grammatical/lexical aspects" while leaving untapped the potential for "integrated use of immersive technologies and participatory storytelling for linguistic heritage revitalisation through active community engagement". This documentation-oriented paradigm, while valuable, risks treating language as a static artefact rather than a living, evolving practice intimately bound to identity and culture.

The Pangloss Collection exemplifies the documentation paradigm at its most sophisticated. Hosting 1,180 hours of audio and video recordings of approximately 200 languages, it represents a significant contribution to the long-term archiving of endangered linguistic heritage. Yet even as the collection provides open access through a bilingual interface, it raises questions about how such archives function as living resources for the communities whose languages they document. The challenge, as scholars have noted, is that "memory, tradition, and heritage cannot be saved, only rematerialized, renegotiated, and reinvented" through performative acts of cultural engagement.

2.3 Language, Literature, and Identity

The relationship between language, literature, and identity is deeply constitutive. Literature—whether oral or written—serves as both a mirror and a shaper of collective identity, encoding cultural values, historical experiences, and epistemological frameworks. Vernacular literary traditions resist homogenisation and erasure, functioning as "a vital marker of identity and cultural continuity". In the context of endangered languages, literature takes on heightened significance as both a repository of linguistic knowledge and a site of identity affirmation.

Recent scholarship has emphasised the multimodal nature of indigenous language practices. Research on "Multimodal Language Rights and Indigenous Totemic Identity" demonstrates that indigenous language practices encompass digital storytelling, singing, e-book making, dance, and video making—forms that mainstream education systems often fail to recognise or validate. This finding has significant implications for digital preservation initiatives, suggesting that effective interventions must engage with the full spectrum of multimodal cultural expression rather than privileging text-based paradigms.

2.4 Digital Storytelling and Cultural Sustainability

Digital storytelling has emerged as a particularly promising methodology for cultural sustainability. Studies of museum innovation in Southeast Asia demonstrate that digital storytelling technologies can transform cultural exhibits into interactive and sustainable experiences, increasing visitor engagement while reducing environmental impact. More fundamentally, digital storytelling enables communities to narrate their own histories and identities, resisting the extractive research methodologies that have historically characterised anthropological engagement with indigenous communities.

The REVIVE project represents a significant advance in this area, deploying Extended Reality (XR) technologies to create an online digital scaffolder—REVIVE(S)—that integrates historical testimonies with new XR digital narratives. The project's dual focus on Cornish, a revived language, and Griko, a severely endangered language, demonstrates the applicability of digital storytelling across different contexts of endangerment.

2.5 The Ethics of Digital Preservation

As digital preservation initiatives proliferate, ethical questions have come to the fore. The UNESCO LINKS programme emphasises that "technology must be used ethically, with clear respect for ownership and cultural protocols", and that "true preservation depends on community control and participation". This principle of community data sovereignty challenges conventional archival paradigms, asserting that indigenous peoples must "remain owners of their data, knowledge, and cultural heritage".

The concept of Linguistic Data Sovereignty has gained traction as a framework for ensuring that digital preservation efforts do not replicate colonial extractive practices. AI-driven language revitalisation tools must function "not as an external savior but as a collaborative epistemic ally that empowers linguistic communities". This requires what scholars describe as "community-centred" approaches to AI development, ensuring that technology "respects linguistic diversity and preserves cultural authenticity".

3. Methodology

3.1 Research Design

This study employs a qualitative, multi-case study research design. The case study approach is particularly suited to investigating complex, context-dependent phenomena such as digital cultural preservation, where the boundaries between phenomenon and context are inherently blurred. The research adopts an interpretivist epistemological stance, recognising

that cultural preservation is not a neutral technical activity but a value-laden social practice shaped by power relations, historical contingencies, and community aspirations.

3.2 Case Selection

Cases were selected through purposive sampling to ensure diversity across geographic, linguistic, and technological dimensions. The selection criteria included: (1) explicit focus on endangered or marginalised languages; (2) integration of digital technology for preservation or revitalisation; (3) evidence of community engagement or co-creation; and (4) availability of documented outcomes. The following cases were selected:

1. **REVIVE Project** (European Union): Focuses on Cornish (UK) and Griko (Italy), deploying XR technologies and participatory storytelling for linguistic heritage revitalisation.
2. **Pangloss Collection** (France): A digital archive hosting 1,180 hours of recordings from approximately 200 endangered languages.
3. **BHASHINI Platform** (India): A national AI-powered language translation platform supporting 22 languages (voice) and 36 languages (text) with over 350 AI models.
4. **Collaborative Digital Heritage Studio** (CoDHerS) (Canada): A community-first digital heritage studio working with Indigenous and minoritised communities.
5. **Bharatvani Project** (India): A digital archiving initiative for Indian languages, particularly those with limited digital resources.

3.3 Data Collection

Data were collected from multiple sources: (1) project documentation and reports; (2) peer-reviewed academic publications; (3) institutional websites and policy documents; (4) UNESCO and government publications; and (5) media coverage and expert commentary. This triangulation of sources enhances the credibility and robustness of findings.

3.4 Analytical Approach

Thematic analysis was employed to identify patterns and themes across the case studies. The analysis proceeded through the following stages: (1) familiarisation with data; (2) generation of initial codes; (3) searching for themes; (4) reviewing themes; (5) defining and naming themes; and (6) producing the report. The analysis was guided by the research question and informed by the conceptual framework of "heritage-to-horizon" sustainability.

4. Case Studies

4.1 REVIVE: XR and Participatory Storytelling for Endangered Languages

The REVIVE project (Community-centred language REVItalisation powered by heritage-led storytelling and digital innoVation in Europe) is a €3 million initiative funded by Horizon Europe and led by the StoryLab Research Institute at Anglia Ruskin University. Bringing together partners from the UK, Italy, Germany, Portugal, Romania, Serbia, and Estonia, the project addresses the challenge that "linguistic and cultural diversity is part of Europe's shared heritage, but many languages are disappearing, particularly in vulnerable ecological and social contexts".

The project's innovation lies in its integration of Extended Reality (XR) technologies with participatory storytelling. Rather than merely documenting grammatical structures, REVIVE produces "the first EU online digital scaffolder—so called REVIVE(S)—comprising a) an online repository (and showcase) of old material testimonies collected by locals and new XR digital narratives and b) onsite temporary exhibitions". The project operates through two case studies: Cornish, a language that had no surviving native speakers in 1800 but has since been revived, and Griko, a severely endangered language spoken in southern Italy.

A critical dimension of REVIVE is its emphasis on community-centred approaches. The project recognises that "although digital technologies provide a uniquely powerful opportunity to revitalise EIHL, previous efforts have predominantly concentrated on preserving grammatical/lexical aspects" and that there remains "untapped potential in exploring the integrated use of immersive technologies and participatory storytelling for linguistic heritage revitalisation through active community engagement". The project's objective is to demonstrate that "citizen participation, multidisciplinarity, and the creative connotation of digital innovation are crucial and effective means for consolidating and promoting multilingualism".

4.2 Pangloss Collection: A Digital Archive of Linguistic Diversity

The Pangloss Collection, developed by the French National Center for Scientific Research (CNRS) since the 1990s, represents one of the most ambitious digital archiving initiatives for endangered languages. The collection "hosts 1,180 hours of audio and video recordings of about 200 languages," with materials "archived for the long-term using a suite of French public services for digital humanities". Accessible through a bilingual English-French website, the collection offers both a general-audience interface and a professional interface, democratising access to linguistic data while maintaining scholarly rigour.

The Pangloss Collection raises important questions about the sustainability of digital archives. As one study of digital obsolescence notes, "while technological impermanence is intrinsic to the digital domain, it poses serious risks to the accessibility and long-term preservation of humanities data". For endangered language archives, this risk is particularly acute: "the collected and archived data on endangered languages may represent the only record of their kind available to future generations. In the case of critically endangered languages, it may constitute the sole remaining documentation of the language itself".

The collection's long-term sustainability depends on continuous format migration and adherence to standards such as the FAIR principles (Findable, Accessible, Interoperable, Reusable). The experience of the Atlante Lessicale Toscano (ALT) demonstrates both the challenges and opportunities: "painstakingly collected linguistic data can be transformed into sustainable, interoperable formats that ensure long-term usability and scholarly value".

4.3 BHASHINI: AI-Powered Language Preservation in India

India's linguistic landscape is extraordinarily diverse, with 22 scheduled languages and hundreds of non-scheduled languages spoken by minority communities. The BHASHINI platform, launched in 2022, represents a significant national effort to harness artificial intelligence for language preservation and accessibility. The platform supports "22 languages (voice) and 36 languages (text), with 350+ AI models and over 4 billion+ language transactions".

BHASHINI operates as part of a broader governmental strategy to integrate AI within cultural and linguistic preservation. The BhashaDaan initiative, conducted under the Ministry of Minority Affairs, is a "citizen-driven initiative under BHASHINI" that engages communities in contributing linguistic data. Complementary projects include the "Kumaoni Voice" project and the "Sora" language documentation initiative, as well as AI-powered projects to revitalise the Rabha and Tiwa languages of Assam.

The BHASHINI model demonstrates the potential of national-scale infrastructure for language preservation. However, it also raises questions about the relationship between state-led initiatives and community autonomy. The integration of AI into language preservation requires careful attention to data sovereignty, as communities must retain ownership and control over their linguistic heritage. As one study cautions, AI must function "as a collaborative epistemic ally that empowers linguistic communities" rather than as an external solution imposed from above.

4.4 Collaborative Digital Heritage Studio (CoDHerS): Community-First Digital Preservation

The Collaborative Digital Heritage Studio (CoDHerS) at the University of British Columbia exemplifies the community-first approach to digital heritage preservation. Founded by Dr. Aynur Kadir, the studio works with Indigenous and minoritised communities "to document and strengthen endangered languages and cultural knowledge through interactive media, digital archives, and multimedia tools".

CoDHerS's approach is explicitly designed to counter the extractive research methodologies that have historically characterised anthropological engagement with indigenous communities. As Kadir explains, "Anthropology, in particular, has a long history of extractive research methodologies tied to colonialism, taking traditional knowledge to advance academic careers without directly working with community needs. Our goal is to change that". The studio's community-first philosophy ensures that "community members needed to be at the driver's seat" of collaborative projects.

The studio's projects demonstrate the transformative potential of community-led digital preservation. The Indigenous Languages Typing Game Initiative, which supports keyboard lessons in both Uyghur and Māori, emerged from a child's simple question about why typing games were only available in English. The "Languages Carry Worlds: Language Preservation" project, a student-led initiative, reflects on relationships to language, memory, and identity.

CoDHerS provides a full suite of audio recording and podcasting tools, DSLR cameras, mobile filmmaking setups, 3D scanners, and 360-degree video cameras, which can be brought to communities locally or internationally. This infrastructure enables small, community-driven ideas to grow into meaningful cultural projects, demonstrating that technology is most effective when it serves community-defined goals rather than externally imposed agendas.

4.5 Bharatvani: Digital Archiving for Indian Languages

The Bharatvani project, a systematic digital archiving initiative for Indian languages, addresses the critical need for preserving linguistic diversity in one of the world's most linguistically rich nations. A systematic review of the project highlights its significance in the broader context of Indian language preservation, where many languages face marginalisation due to the dominance of official and regional languages.

The project's approach to digital archiving recognises that preservation must extend beyond documentation to include revitalisation. By creating accessible digital resources, Bharatvani aims to support language learning, research, and cultural engagement. The project's

systematic review methodology provides valuable insights into the challenges and opportunities of large-scale digital archiving in linguistically diverse contexts.

5. Findings and Discussion

5.1 Technology as Enabler, Not Solution

A consistent finding across all case studies is that technology functions as an enabler rather than a solution in itself. The mere digitisation of linguistic and literary materials does not guarantee their preservation or revitalisation. As the REVIVE project demonstrates, "digital technologies provide a uniquely powerful opportunity to revitalise EIHL," but this potential is realised only through "active community engagement" and "participatory storytelling". Similarly, the CoDHerS studio's experience shows that technology must be deployed in service of community-defined goals, with "community members at the driver's seat" of collaborative projects.

This finding has significant implications for policy and practice. Investment in digital infrastructure must be accompanied by investment in community capacity-building, digital literacy, and participatory design processes. The digital divide cannot be bridged by technology alone; it requires attention to the social, economic, and political conditions that shape access and participation.

5.2 Community Participation as Prerequisite for Sustainability

The case studies converge on the finding that community participation is not merely desirable but is a prerequisite for sustainable digital preservation. The REVIVE project's emphasis on "citizen participation, multidisciplinarity, and the creative connotation of digital innovation", the CoDHerS studio's community-first philosophy, and UNESCO's insistence that "true preservation depends on community control and participation" all point to the same conclusion: sustainable preservation is participatory preservation.

This finding challenges conventional top-down models of heritage preservation. It suggests that the most effective digital preservation initiatives are those that begin with community needs and aspirations, that engage community members as co-creators rather than passive subjects, and that ensure community ownership of data and cultural heritage. The principle of Linguistic Data Sovereignty provides a useful framework for operationalising this commitment.

5.3 The Multimodal Nature of Linguistic Heritage

The case studies reveal that linguistic heritage is inherently multimodal, encompassing not only written and spoken language but also visual, performative, and material forms of

expression. Research on "Multimodal Language Rights and Indigenous Totemic Identity" demonstrates that indigenous language practices include digital storytelling, singing, e-book making, dance, and video making. Effective digital preservation must therefore engage with the full spectrum of multimodal cultural expression rather than privileging text-based paradigms.

This finding has implications for the design of digital preservation platforms. Such platforms must be capable of accommodating diverse media formats, supporting interactive and immersive experiences, and enabling communities to represent their heritage in culturally appropriate ways. The REVIVE project's use of XR technologies and the CoDHerS studio's investment in 360-degree video and 3D scanning equipment represent significant advances in this direction.

5.4 Ethical Governance and Data Sovereignty

The case studies highlight the critical importance of ethical governance and data sovereignty in digital preservation. As UNESCO emphasises, "technology must be used ethically, with clear respect for ownership and cultural protocols". This requires that communities "remain owners of their data, knowledge, and cultural heritage". The principle of Linguistic Data Sovereignty provides a framework for ensuring that digital preservation efforts do not replicate colonial extractive practices.

The challenge of data sovereignty is particularly acute in the context of AI-powered language preservation. While AI tools offer powerful capabilities for documentation, translation, and language learning, they also raise concerns about the commodification of endangered languages and the appropriation of indigenous knowledge. AI must function "not as an external savior but as a collaborative epistemic ally that empowers linguistic communities", with communities retaining control over how their linguistic data is used.

5.5 From Documentation to Revitalisation: A Paradigm Shift

The case studies reveal a paradigm shift in digital preservation, moving from documentation-oriented approaches to revitalisation-oriented approaches. Traditional preservation focused on recording grammatical structures, vocabularies, and texts as static artefacts. Contemporary approaches, exemplified by REVIVE and CoDHerS, position digital preservation as a dynamic process of cultural engagement, creativity, and identity re-articulation.

This paradigm shift is reflected in the language used to describe preservation initiatives. Terms such as "revitalisation," "reclamation," and "sustainability" have largely replaced "documentation" and "archiving," reflecting a new understanding of what digital preservation

can and should achieve. The goal is not merely to preserve language for future study but to sustain language as a living practice of cultural identity and expression.

5.6 A Conceptual Framework: Heritage-to-Horizon Sustainability

Synthesising the findings from the case studies, this paper proposes a conceptual framework for "heritage-to-horizon" sustainability in digital preservation. The framework consists of four interconnected dimensions:

1. **Cultural Continuity:** Digital preservation must contribute to the intergenerational transmission of linguistic and literary heritage, ensuring that cultural knowledge is not merely archived but actively shared and practised.
2. **Community Agency:** Preservation initiatives must be community-led, with community members exercising control over data, decision-making, and the representation of their heritage.
3. **Technological Innovation:** Digital technologies must be deployed creatively and ethically, in ways that serve community-defined goals and respect cultural protocols.
4. **Environmental Sustainability:** Digital preservation must be environmentally sustainable, minimising energy consumption and material use throughout the lifecycle of digital resources.

This framework positions digital preservation not as a backward-looking exercise but as a forward-oriented practice of cultural innovation. The "horizon" of the conference theme becomes the space of possibility where heritage and innovation converge, where the past informs the future, and where language, literature, and identity are sustained as living resources for generations to come.

6. Conclusion

This paper has examined the intersection of language, literature, and identity in the digital preservation of regional cultural narratives. Through a multi-case study analysis of initiatives spanning Europe, India, and Canada, it has identified key findings that contribute to both academic understanding and practical policy.

The study's central conclusion is that digital preservation is most effective when it is community-driven, ethically governed, and oriented towards revitalisation rather than mere documentation. Technology is a powerful enabler but not a solution in itself; the sustainability of digital preservation depends on the social and cultural conditions in which it is embedded.

The proposed "heritage-to-horizon" framework offers a conceptual contribution to the field, synthesising insights from the case studies into a coherent approach that bridges cultural

heritage and future-oriented innovation. This framework has implications for policymakers, practitioners, and communities engaged in digital preservation initiatives.

For policymakers, the study recommends: (1) investment in community capacity-building and digital literacy; (2) development of legal and ethical frameworks that ensure Linguistic Data Sovereignty; (3) support for multimodal and immersive approaches to digital preservation; and (4) integration of digital preservation into broader sustainable development strategies.

For practitioners, the study emphasises the importance of participatory design, cultural appropriateness, and long-term sustainability planning. Digital preservation initiatives should be designed with, not for, communities; they should engage the full spectrum of multimodal cultural expression; and they should anticipate the challenges of technological obsolescence and resource sustainability.

For communities, the study affirms the value of local knowledge, creativity, and agency in shaping the digital futures of their linguistic and literary heritage. The tools of digital preservation are most powerful when they are in the hands of those whose heritage they are designed to sustain. As the world confronts the accelerating loss of linguistic and cultural diversity, the stakes of digital preservation could not be higher. The challenge is not merely to archive the past but to sustain the conditions for cultural flourishing in the future. This is the promise and the imperative of "heritage-to-horizon" sustainability.

References

- Adamou, E., Guillaume, S., & Michaud, A. (2026). The Pangloss Collection: Opening up research data on endangered and underdocumented languages. Language.
- Cucurullo, S., Montemagni, S., Saccoccio, R., & Sassolini, E. (2025). The challenge of obsolescence of digital archives in cultural heritage. IEEE Xplore.
- Dayanand, A., & Sundaram, U. D. (2026). Reviving Indigenous voices: Digital storytelling and mobile-assisted language learning for the Irula community. In Digital Storytelling and Mobile-Assisted Language Learning (pp. 145-162). Taylor & Francis.
- Kadir, A. (2026). In community hands: Sustaining language and culture in a digital world. UBC Faculty of Arts.
- Meilani, M. (2025). Digital storytelling for sustainable cultural heritage: Technology-driven museum innovation in Southeast Asia. IOP Conference Series: Earth and Environmental Science, 1564, 012108.

- Mills, K. A., Doyle, K., & Friend, L. (2025). Multimodal language rights and Indigenous totemic identity: Cross-cultural participatory research. Journal of Language, Identity, and Education, 24(2), 449-467.
- Saxena, S., & Kaur, G. (2025). Preservation and revitalisation of Indian languages through digital archiving: A systematic review of Bharatvani. DESIDOC Journal of Library & Information Technology.
- UNESCO. (2026). Indigenous knowledge in the digital age: Learning, preserving, sharing.
- UNESCO. (2026). International Decade of Indigenous Languages (2022-2032).
- Zaidi, N., & Kasana, S. (2025). Vernacular encounters: Politics and possibilities. Taylor & Francis.