Community Archiving of Ethnic Groups in Thailand

Siripen Ungsitipoonporn

Research Institute for Languages and Cultures of Asia,
Mahidol University, Thailand

Buachut Watyam

Research Institute for Languages and Cultures of Asia,
Mahidol University, Thailand

Vera Ferreira

ELAR, SOAS, University of London, UK

Mandana Seyfeddinipur

ELAR, SOAS, University of London, UK

This article presents the research process of the project "The Ethnic Group Digital Archive Project: Promoting the protection and preservation of language and culture diversity in Thailand". This project involved the development of a local digital archive website for the ethnic groups of Thailand to archive, preserve, and transmit their knowledge of languages and cultures to their younger generations and those interested. The core objective of this digital archive development was the implementation of the archive website with uncomplicated accessibility and simple and interesting design that serves the language documentation purpose. The digital archive output includes collections from 18 ethnic groups in Thailand, containing 385 bundles of legacy and fieldwork data obtained by means of video, audio, text, image, and ELAN file. Despite the low number of researchers working on language documentation and archiving, the research team managed to expand both national and international networks working in this particular field of study. This serves as an opportunity for scholars and speaker communities in Thailand to recognize the importance of local knowledge preservation and transmission, and the availability of the digital archive is a practical way to support sustainable data preservation and accessibility in the future.

1. Introduction[1] Thailand has more than 70 living languages, belonging to five language families: Tai (24 languages), Austroasiatic (23 languages), Austronesian (3 languages), Sino-Tibetan (18 languages), and Hmong-Mien (2 languages). The promotion of Thai as the official and national language and its use as medium of instruction in a formal school system puts small languages under threat. At least 15 languages out of 70 are considered to be seriously endangered (Premsrirat et al. 2007; Premsrirat & Burarungrot 2018). Nine of these languages belong to the Austroasiatic language family, which includes the main languages used in mainland Southeast Asia, notably Chong, Kasong, Samre, Chung, Nyah Kur, Lavua, So (Thavueng), Maniq, and Mlabri. Three languages are Tibeto-Burman languages – Gong, Bisu, and Mpi – and two are Austronesian – Urak Lawoc and Moklen. The last, Saek, is a language which belongs to the Tai-Kadai language family (Premsrirat & Burarungrot 2018).

The Research Institute for Languages and Cultures of Asia (RILCA), Mahidol University has been continuously working on the revitalization of endangered languages of ethnic groups in Thailand for more than twenty years, beginning with the Chong of Chanthaburi Province, whose language was at an endangered stage (Premsrirat 2002). At the outset, linguists from Mahidol University mostly conducted descriptive research, but a core subject, "Field method in Linguists", allowed master's degree students the opportunity to work in the field so as to acquaint themselves with community members firsthand. Most dissertations from master's and doctoral students of RILCA have undertaken research on phonological and syntactic description because in the past, many small languages with no written language needed to know their sound system. Therefore, linguistic programs at RILCA encouraged students to do pure linguistics research. This basic research on phonological systems provided a good foundation for language revitalization and preservation. In 2004, the Center for Documentation and Revitalization of Endangered Languages and Cultures was formally established by Professor Suwilai Premsrirat. The first phase of the center aimed at documenting endangered and other indigenous languages of Thailand, and later on, it started working closely with numerous language communities on language development, language revitalization, and community improvement. Linguists from RILCA were able to see how the language situation of ethnic groups is changing due to several factors, such as education, economy, politics, globalization, and interracial marriage, and as a result, the community members, especially in the smaller ethnic groups, have few chances to use their native languages. While linguists compiled language data for phonology, grammar, and discourse analysis with the support of native speakers, the language speakers also needed a tool that they could use to write and record their knowledge and literature. Linguists normally use phonetic symbols as a tool for language documentation and description whereas the language speakers needed scripts that were simple and familiar to them. Orthography development is

one of the most important activities of the language revitalization project and derived from the needs of language speakers (Premsrirat & Burarungrot 2018). It was the Chong people of Chanthaburi Province who became the first group for whom a writing system was developed for their language (Premsrurat 2000; 2002). Soon after, other ethnic groups – such as Nyah Kur, So (Thavueng), Lavue, and Gong – followed the Chong model, and together with linguists, developed their own writing systems using a Thai-based script.

Language and cultural revitalization strategies are different, depending on the context of each community. For example, teaching indigenous languages as a subject in the formal school system is used for Chong (Premsrirat, Ungsitipoonporn, & Choosri 2007; Premsrirat & Rojanakul 2015), Nyah Kur in Chaiyaphum (Ungsitipoonporn 2018), and So (Thavueng) in Sakon Nakhon, and a bilingual education approach is used for Malayu in the deep south (Premsrirat & Burarungrot 2018), Northern Khmer in Surin (Ungsitipoonporn & Laparporn 2019), and Lavue in Mae Hong Son province (Thawornpat 2018). The Nyah Kur language is classified as part of the Monic sub-branch of the Mon-Khmer branch of the Austroasiatic language family (Premsrirat et al. 2004: 19; Ungsitipoonporn 2018: 81). It is also known as an ancient Mon language from the Thavarawadee period (Diffloth 1984). The Nyah Kur people live in three provinces of Thailand – Petchabun, Chaiyaphum, and Nakhon Ratchasima – but the majority of people who speak the Nyah Kur language now live in Chaiyaphum Province (Ungsitipoonporn 2018). The Lavue are an ethnic group that settled in the central Thanon Thong Chai Range in the Chiang Mai and Mae Hong Son Provinces. Their language belongs to the Austroasiatic family, but it was formerly categorized as being in the Palaung-Wa branch of the Mon-Khmer language family (Thawornpat 2018: 143). The Patani Malay language is the first language (or mother tongue) of the people in Thailand's Deep South. This language is different in terms of phonology and syntax from Classical Malay (Jawi script) used in Islamic religious documents and from Standard Malay (Rumi or Romanized-based script) which is used as the national language of Malaysia. These three languages (Patani Malay, Classical Malay, and Standard Malay) have each evolved in their own direction. Patani Malay has been more influenced by Thai, whereas Classical Malay has been impacted by Arabic and Standard Malay by English. Patani Malay is widely used as a spoken language (Premsrirat & Samoh 2012; Samoh 2018). The revitalization project has developed a large number of teaching materials, including big books and small books based on local knowledge found in stories, songs, and conversations. Each type of teaching material has a different purpose and the teaching methods used are also different (Malone & Malone 2013). For example, a booklet is "a small book" that fits in your hand, is suitable for reading alone, and can be easily taken anywhere with you. Small books are used to practice reading, focusing on reading to understand the story content, and are suitable for those who are starting to learn to read-write in any language. Teachers use booklets to encourage students to love reading, writing, and for creative thinking because students can come up with words or events they know to represent events or characters in the story. "Big books" are larger than normal books, around A3 paper size or more. They are used for shared reading among

the whole class. This teaching material emphasizes the meaning of the story and is followed up with grammar learning. A big book contained duplicated sentence patterns – its dominant feature – so that the learners can remember the grammatical structure and predict the content. Both small and big books are used as reading development materials for children, and the content should be based on local culture (Ungsitipoonporn 2018; Ungsitipoonporn & Laparporn 2019). These materials are all handmade, which makes them difficult to reproduce, and they tend to deteriorate or get lost easily. Furthermore, it was hard to collect the data since most of the languages are endangered, and those with knowledge of the language and traditions are few and far between.

There are several well-known digital archives such as ELAR (Endangered Languages Archive), of SOAS University of London;[2] Archives of the Max Planck Society;[3] The Archive of Austroasiatic Intangible Heritage – a digital media resource documenting languages of the Austroasiatic language family of Mainland Southeast Asia and India at Lund University;[4] and Catawba Indian Nation Archives.[5] They have similar objectives, namely, to be resources for the collection and preservation of certain types of data related to particular ethnic groups, be it tangible or intangible data. However, each digital archive has developed methodologies which were specifically tailored in order to provide maximum benefit to their users and contributors.

The advancement of technology today is phenomenal. In 2014, the author had an opportunity to visit SOAS University of London and learned about its digital archive (ELAR), and then in 2016 with funding from Newton Fund and Thailand Research Fund (TRF) for RILCA, Mahidol University, she undertook work in collaboration with ELAR SOAS to develop a digital archive of ethnic groups in Thailand under the project "The Ethnic Group Digital Archive Project: Promoting the protection and preservation of language and culture diversity in Thailand". The main goals of this research project were to convert all legacy data of ethnic groups previously collected in analog format to digital format, publish them online, and develop a management system to establish an archiving standard. This project gave precedence to language consultants who were native speakers, who were all provided access to the data to which they had contributed and that would be useful to other interested parties as well.

Only a small number of digital archives for the languages and cultures of ethnic groups are available in Thailand, and organizations actively working on this matter include Princess Maha Chakri Sirindhorn Anthropology Centre[6] and National Electronics and Computer Technology Center or NECTEC.[7] The Resource Center for Revitalization and Maintenance of Endangered Languages and Cultures at RILCA, Mahidol University, in particular, has been collecting a large amount of data over the years and continues to work collaboratively with several ethnic groups in Thailand

---

[2]https://www.soas.ac.uk/elar/

[3]https://www.archiv-berlin.mpg.de/

[4]https://projekt.ht.lu.se/rwaai

[5]http://catawbaarchives.libraries.wsu.edu/

[6]https://www.sac.or.th/main/

[7]https://www.nectec.or.th/

wishing to maintain language diversity by preserving and revitalizing their languages and cultures.

Theories and practices of language documentation and archiving are constantly being developed in conjunction with the advancement of the technology, aiming at the long-term preservation of endangered languages and cultures. The history of archiving development is worth briefly reviewing here to provide context to the future of digital archiving and its benefit to society.

2. Literature Review

2.1 Review of language documentation and archiving history According to Henke and Berez-Kroeker (2016), several key factors changed traditional archiving of endangered languages, including greater awareness of language endangerment, technological advancement, and the new language research trends.

Henke and Berez-Kroeker (2016) have drawn attention to four periods of archiving development since the late 19th century. The earliest period was when language archives contained mostly paper documents deposited at the archives of universities and museums and people could only access them in person. A huge limitation of the use of paper records is that they deteriorated over time, so data was hard to preserve. Endangered languages attracted more attention during the second period of archiving development early in the twentieth century, when documentation became a practice for endangered language preservation and revitalization. The 1930s to 1980s was a period when the practice involved new technologies that allowed linguists to collect data through audio recording using wax cylinders, phonographs, tape recording, and, ultimately, video. The third period was in the 21st century when language documentation and archiving were challenged and refined by developing better practice standards to achieve "a lasting, multipurpose record of a language" (Himmelmann 2006:1).

However, at that time, language research projects were solely conducted by linguists, and the results mainly served academic purposes and barely benefited the speaker communities themselves. In the current period, eventually though, participatory models were introduced to the language archiving field, the primary factors for which, according to Henke and Berez-Kroeker (2016), were community-oriented research, indigenous community empowerment, and social media. In brief, the participatory models aimed at engagement between the researchers and the target indigenous communities to collaboratively conduct research projects that considered the best interests of both linguists and data contributors. Furthermore, active participation means that communities, especially the smaller speech communities, are empowered with the knowledge and skills to create their own data materials and even their own archives. To clarify, archive websites provide essential components such as protocols for the access to data and materials and a preservation layer that helps mitigate deterioration and facilitates updating all materials in more durable file formats for a long-term preservation. Concurrently, the language archives were

designed to expand potential audience and attract users by disseminating and promoting language documentation and archiving to preserve cultural heritage and diversity sustainably.

2.2 Review of digital archives Digital archives around the world have different purposes, audiences, and users with the best known ones attracting a greater share of academics and scholars such as linguists. Holton (2012) points to two digital archives in particular, the Association of Newfoundland and Labrador Archives (ANLA) and California Language Archive, as examples that have expanded their audiences to include indigenous communities.

ANLA[8] is the archive of Newfoundland and Labrador, the most easterly province of Canada. ANLA aims to preserve archival resources of the province, and it not only provides educational materials but also the most updates to existing archives in Newfoundland and Labrador. The archival catalogue of ANLA can be browsed by description, authority record, institution, function, subject, place, and digital object. Similarly, the California Language Archive[9] is an online archive based at the University of California, Berkeley which provides online resources of physical and digital materials for the purposes of indigenous language documentation and preservation in the Americas. The catalogue can be browsed in several ways, namely, by collection, language, people, or map. These two archives are reported to have native language community members accessing the resources on their languages and cultures on a regular basis, which sets a positive example of the participatory model and the expanded use of archives.

As mentioned in the introduction, there are only a few organizations working on archiving information on ethnic groups in Thailand, including Princess Maha Chakri Sirindhorn Anthropology Centre (SAC) and National Electronics and Computer Technology Center (NECTEC).

Princess Maha Chakri Sirindhorn Anthropology Centre is a center under the supervision of the Ministry of Culture of the Kingdom of Thailand. It is basically a database for the anthropology of Thailand, archiving several categories of data such as ethnic groups in Thailand, rituals, ceremonies, rare books from the Siam era, folk toys, and inscriptions. In the Ethnic Groups of Thailand database, for example, SAC has mostly general information on 70 ethnic groups, including population data, history, traditions, customs, religions, and so on. NECTEC is a government organization under the National Science and Technology Development Agency of the Ministry of Higher Education, Science, Research, and Innovation. It has developed a platform named Navanurak[10] to preserve arts and cultures in digital format. Navanurak is a digital archive for storing cultural data and media specifically related to the artistic and cultural objects of local museums throughout Thailand. The platform allows trained and verified users, to collect museum data and examine displayed objects

---

[8]http://anla.nf.ca/
[9]https://cla.berkeley.edu/
[10]http://www.navanurak.in.th/

themselves by providing manuals to explain the process of archiving from start to finish.

The main aim of this article is to present information about the implementation of a local digital archive website named LangArchive-TH[11] and to describe past activities and the accomplishments of the research project, including the digitization process of analog materials, visits to ethnic communities to test the collected digital data, and the language documentation and archiving workshops held for scholars working with the ethnic communities. The results of the project are exhibited in a number of collections and digital heritage items archived on LangArchive-TH, as well as an account of the obstacles that arose during the project, suggestions for archive improvement, and future plans.

3. Objectives

- To present the implementation of LangArchive-TH

- To present the project's activities and deliverables

- To present the digital archive statistics and dissemination of results: the evaluation, feedback and recommendations

4. Research methodology   The methodology of this project consists of three core parts as follows:

1. The implementation of the local digital archive using Mukurtu CMS

2. Data management and archiving

3. The digital archive website dissemination

4.1 The implementation of the digital archive website using Mukurtu CMS   The research project aimed for long-term ethnic language and cultural data preservation and the dissemination of information about the digital archive of languages and cultures of ethnic groups of Thailand. To achieve this objective, the community-oriented digital archive LangArchive-TH was created and implemented on Mukurtu CMS, which is a content management system (CMS) tailored for community use and designed to be user-friendly and easy to access. This way, the project hopes to encourage community members to collect and archive their own language and cultural data.

The LangArchive-TH, unlike other academic archives, attracts audiences with its simple design that offers more pictures with less text and fewer menus. As the goal was to empower researchers and community members in the use of archived materials for research and revitalization purposes, the archive structure was specifically developed and adjusted to suit its main audience in Thailand. To do this, the web interface was localized and translated from English into Thai. Even though LangArchive-TH is

---

[11]https://langarchive-th.org/

designed to be user-friendly and provide easy access to digital archive websites, security is not left wanting. The Mukurtu CMS provides options to manage communities and protocols that can specify audiences for each bundle of data or digital heritage item to meet the contributors' requirements.

4.2 Data management and archiving The ethnic group digital archive project exploited language and cultural data of ethnic groups in Thailand from two sources: legacy data and fieldwork data.

4.2.1 Legacy data Legacy data refers to data collected over the past 20 years by the Resource Center for Revitalization and Maintenance of Endangered Languages and Cultures, Research Institute for Languages and Cultures of Asia, Mahidol University. Most of the legacy material was in an analog format, e.g., cassettes tapes, video tapes, reversal films and negative films. The materials needed to be stored in the communities and at the resource center as some of them had gotten lost or had severely deteriorated. In order to preserve the language and cultural knowledge of ethnic groups in Thailand, the legacy data needed to be digitized to be effectively managed and uploaded to an archive available to the general public to foster academic and community-driven use of the archived material.

The research team from RILCA undertook a training workshop entitled "Digitization, Data Management & Archiving" together with the research team from ELAR at SOAS University of London. The workshop focused on digitization of legacy material, data management and curation, and data archiving. The RILCA team were introduced to the digitization process for different types of analog media, e.g., converting cassette tapes to MP3 format with a cassette tape player and converting films to JPG format with a film scanner. The digitization process was digitized to illustrate the precise sequence and compiled into a manual for use by the team in future archiving endeavors.

4.2.2 Fieldwork data The fieldwork data samples were collected in high-quality digital format by the RILCA research team during field visits to ethnic communities whose native languages were at an endangered stage. The research team visited Chung community in Kanchanaburi Province and Saek community in Nakhon Phanom Province to collect data on their language and culture. The Chung and Saek languages are both identified as endangered; however, their language situation and contexts are totally different.

The fieldwork for this language documentation and archiving project gave precedence to severely endangered ethnic groups such as the Chung, with its very small population and critically endangered language. Chung is a Pearic language, a branch of the Austroasiatic language family that has approximately 50 speakers left (Premsrirat et al. 2004: 84–85). There are only a few studies on the Chung language, one such being "Investigating contact-induced language change: cases of Chung (Saoch) in Thailand and Cambodia" (Choosri 2007). The language and cultural data of Chung

had never been recorded in a format that was preservable. Based on the Chung field-work, few people could speak their native language, and it was rarely used anymore in daily life. For this reason, a field trip was arranged to the Chung community in Kanchanaburi Province with the aim of updating the language situation.

The Chung language is severely endangered due to the fact that it has not been spoken on a regular basis since the government forced them to live in locations near other ethnic groups such as the Khmu. Few fluent speakers remain in the village, and their unique cultural practices are forgotten. This created a significant obstacle to the process of language documentation, and the research team could only record some custom performances but struggled to elicit natural data. RILCA and ELAR research teams had to consult about data collection methods in order to elicit more natural data. The research team selected picture books as stimuli to encourage the language consultants to tell stories based on the pictures in their own language. Eventually, they were able to increasingly recall the language as they became more familiar with the data collection process.

Saek belongs to the Tai-Kadai language family, and there were 3,500 speakers in Thailand (Premsrirat et al. 2004: 65). Saek has a large number of fluent speakers who were aware of the extinction risk to their language and culture due to there being no orthography with which to write down their local stories or knowledge. While the middle generation were able to speak Saek fluently and still used it in daily life, the younger generation barely spoke the language. Nevertheless, community members had been continuously working on revitalization projects with RILCA, Mahidol university. For example, the community-based research at Bawa District, Nakhon Phanom Province developed a Saek orthography to document their narratives, songs and stories using their own writing system. The community members and the research team agreed to develop the orthography based on Thai script. They worked deliberately on the alphabet development process, including adding diacritics such as low tone marks to the letters of Thai alphabet in order to represent their unique voice. Since most of the members already had Thai language literacy, it was easier to connect knowledge between languages. Hence, they have been able to successfully transmit their local knowledge, customs, and traditions to the younger generation. Even so, the data was not yet recorded on any type of digital media, just written down. For this reason, the research team decided to promote language preservation by enriching existing data such as games, songs, narratives, and cooking procedures by recording them on high quality audio and video for archiving and dissemination purposes. Subsequently, community members were contacted, and the process of language documentation was explained to them. The community members were not ready to conduct the whole process of documentation to archiving themselves as they lacked the digital equipment and technological skills needed, but they were enthusiastic about participating. The research team was eventually able to collect eleven videos featuring cooking, cultural activities, games, and narratives.

All of data collected was systematically organized and archived. The workflow of data management and archiving can be seen in Figure 1.
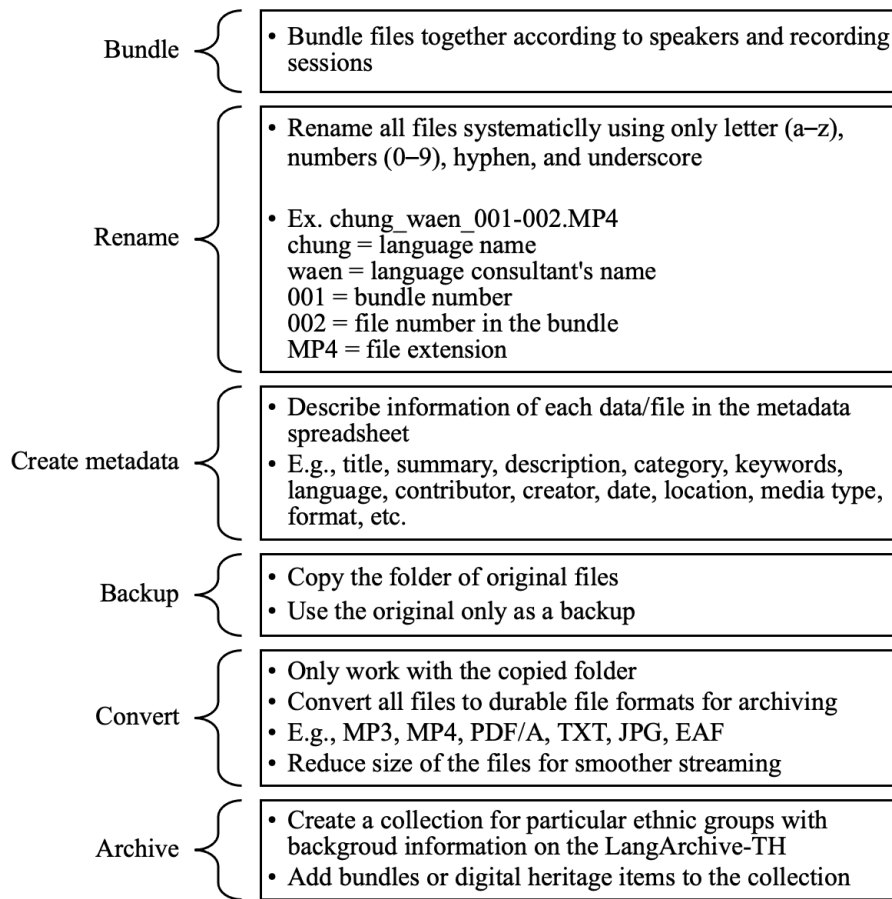
Bundle	• Bundle files together according to speakers and recording sessions
Rename	• Rename all files systematiclly using only letter (a–z), numbers (0–9), hyphen, and underscore • Ex. chung_waen_001-002.MP4 chung = language name waen = language consultant's name 001 = bundle number 002 = file number in the bundle MP4 = file extension
Create metadata	• Describe information of each data/file in the metadata spreadsheet • E.g., title, summary, description, category, keywords, language, contributor, creator, date, location, media type, format, etc.
Backup	• Copy the folder of original files • Use the original only as a backup
Convert	• Only work with the copied folder • Convert all files to durable file formats for archiving • E.g., MP3, MP4, PDF/A, TXT, JPG, EAF • Reduce size of the files for smoother streaming
Archive	• Create a collection for particular ethnic groups with backgroud information on the LangArchive-TH • Add bundles or digital heritage items to the collection

Figure 1. The workflow of data management and archiving

As mentioned, Mukurtu CMS provided the community and protocol management system, and LangArchive-TH offered full authority to depositors to control data assessment by specifying community and protocol since all data rights are reserved to the speech community. In order to protect data contributors and depositors, they retrained full rights to determine whether the data was openly available, available only upon request, or restricted to particular groups of users to access and download the materials. In addition, contributors in the field were involved in the data collection process, especially regarding their decisions of communicative events they wanted to document.

4.3 Dissemination of archive information The dissemination of knowledge gained from the ethnic group digital archive project is supported by information promoting the digital archive website LangArchive-TH to general audiences and disseminating the knowledge concerning language documentation and digital archiving for students, scholars, and university lecturers at the local, national, and international levels.

4.3.1 Local dissemination   Local dissemination was done through workshops, training participants in language documentation, data management, and archiving. Specifically, the workshops aimed to pass on knowledge about language documentation techniques, data management methods, and archiving for local communities to be able to promote the protection and preservation of language and cultural diversity in Thailand. Therefore, the research team mainly worked collaboratively with educational institutions and community members.

The first workshop was held at RILCA, Mahidol University; the ELAR researcher was in charge of the training. One of the main goals of this workshop was to train the trainers. In other words, the RILCA research team, who were the administrators of LangArchive-TH digital archive website, participated in the workshops to gain more experience in digital archiving to improve their competency to effectively conduct the whole process from language documentation to archiving. Besides that, what they learned at this workshop also gave them the capacity to teach and hold future local workshops in Thailand themselves. This workshop was also open to anyone else who was interested. The participants included interested students and scholars from central and provincial regions, as well as So Kusuman native speakers from Sakhon Nakhon Province.

The RILCA research team successfully held three more workshops in Trat, Pattani, and Chiang Rai provinces, the participants of which were mainly lecturers from educational institutions, namely, Trat Community College, Yala Rajabhat University, Prince of Songkla University, Hatyai University, Kanarat Bamrung Pathum Thani School, and Chiang Rai Rajabhat University. All those who took part enthusiastically engaged in workshop activities, learning and practicing all processes from start to finish.

The workshops in Trat and Chiang Rai were very similar, as in both cases students had opportunities to work with native speakers from nearby ethnic communities and could be involved in all aspects, from collecting and managing data to archiving.

On the other hand, the workshop in Pattani had the limitation of being located in Thailand's restive deep south, so documentation of the languages of the region was not possible. However, most of the participants already had their own data on the Pattani Malay community in analog format, meaning that their training consisted of learning the process of digitization of legacy data, data management, and uploading to the archive.

4.3.2 National dissemination   The Research Institute for Languages and Cultures of Asia, RILCA, Mahidol University in collaboration with the Princess Maha Chakri Sirindhorn Anthropology Centre (SAC – Bank of Anthropological Data), the Thai Film Archive, and Thailand's National Electronics and Computer Technology Center (NECTEC) held a forum under the name "Digital Archives of Thailand: Management and Archiving, and the Creation and Presentation of Digital Archives" at which different archives in Thailand were presented. The main goal of this forum was to promote archive sharing and to enhance the discussion and dissemination of information about digital archive creation and management. With its primary aim being to fos-

ter academic and community-driven use of archived materials, LangArchive-TH was presented to the general public for the first time at this forum. The Q&A section from this forum was published in the book การจัดทำคลังข้อมูลดิจิทัลด้านภาษาและวัฒนธรรมของ กลุ่มชาติพันธุ์ในประเทศไทย (*The digital archive for the languages and cultures of the ethnic groups of Thailand*[12]). The section contained very useful discussions on digital archiving in Thailand and its future.

4.3.3 International dissemination   The digital archive website LangArchive-TH was internationally launched at the international conference on Language Diversity and Preservation of Cultural Heritage (LCP 2019) hosted by RILCA, Mahidol University and organized in collaboration with ELAR SOAS, University of London. The project researchers from RILCA and ELAR presented background information on the joint project of the ethnic group digital archive, Mukurtu CMS platform, and the LangArchive-TH showcase, including data statistics on the archive website.

5. Results   The results of the research include LangArchive-TH website data information and dissemination results – evaluations, feedback, and recommendations. The LangArchive-TH contents include the number of collections, digital heritage items, and media types uploaded to the archive website, as can be seen in the following table.

Table 1. The results of data information of LangArchive-TH

LangArchive-TH numbers				
18 collections: 18 ethnic groups Hakka, Tak Bai, Gong, Pattani Malay, Chung, Saek, Chong, Urak Lawoc, Northern Khmer, Kasong, Nyah Kur, Kuy, Moken, Akha, Bisu, Plang, Lavue, and Khmu				
385 bundles				
Video	Audio	Text	Image	ELAN
226 files	135 files	148 files	360 files	90 files

According to the table, LangArchive-TH currently has data collections related to 18 ethnic groups in Thailand. In total, there are 385 bundles of data available on five types of digital media: video, audio, text, image, and ELAN file. The eighteen ethnic groups involved are categorized based on the language families as listed below.

- Austroasiatic languages: Chung, Chong, Kasong, Nyah Kur, Kuy, Northern Khmer, Khmu, Plang, and Lavue

- Austronesian languages: Pattani Malay, Urak Lawoc, and Moken

- Tai-Kadai languages: Tak Bai and Saek

---

[12]https://langarchive-th.org/th/system/files/atoms/file/Archiving_book.pdf

• Sino-Tibetan: Gong, Bisu, Akha, and Hakka

In the future, LangArchive-TH aims to enrich existing collections with more digital heritage items as well as the collections of other ethnic groups which have yet to be archived. Hence, the archive will provide archival data on all five language families in Thailand. For this matter, the research team had been discussing plans to support larger numbers of recordings in the future. First of all, both physical and online storage were increased to cope with increasing data. Furthermore, one of the most significant things to prepare was the structure of the archive to support the upcoming bundles and collections. To enhance the web browsing experience, LangArchive-TH is now providing sub-collections within each collection by distinguishing academic works and separating them from the data collected for further academic analysis, community-based recordings, and information contributed by the speech communities.

Concerning the evaluation and recommendations, a questionnaire focused on the review of the digital archive website asked the workshop participants to rate and give their opinions about the website after navigating through it. In total, 63 participants completed the workshops, most of whom were lecturers, master's and doctoral students, and members of the ethnic groups.

Figure 2. The home page of www.langarchive-th.org

The questionnaire consisted of two types of questions: rating scale and open-ended questions. The rating scale questions were categorized into six topics. The evaluation result of each topic is shown in Figure 3.

As for the open-ended questions, they focused on the problems the participants encountered while navigating through the archive, browsing data, and viewing or playing media. A common problem was the slow speed, which might have been down to the web host or internet connection. The participants were mostly satisfied with the design, user-friendliness, and data accessibility, given that it was simple and
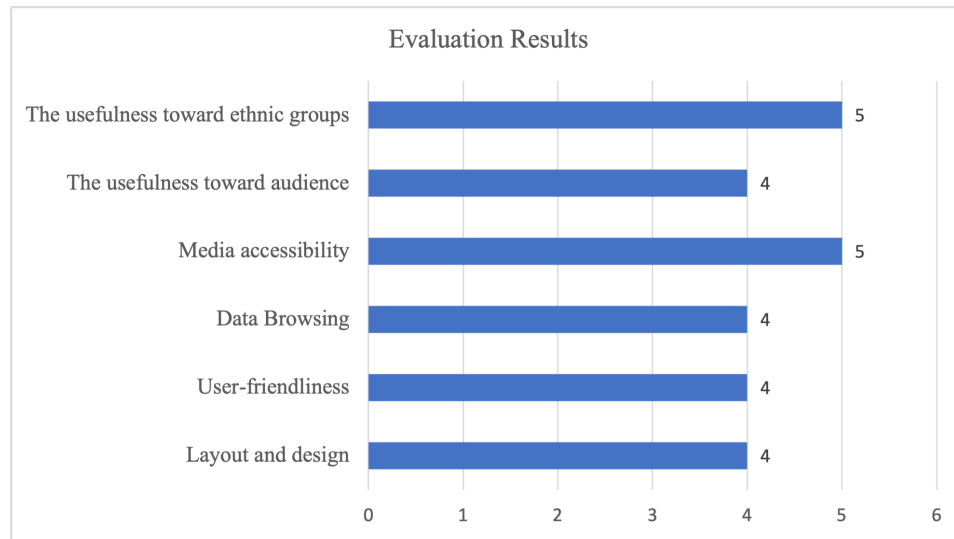
Figure 3. The evaluation results: 1 means Poor, 5 means Very Good

easy to navigate through the website. Even so, some of them found it difficult to browse for data as they needed more time to learn how to search for specific content.

With regard to feedback on the data management and archiving training, most participants commented that the processes of managing and archiving data was not too complicated and in fact easy to learn. It was however more difficult for those who did not possess computer skills; hence they could do with additional training hours in the future. Furthermore, the participants realized that the digital archive for the languages and cultures of the ethnic groups of Thailand would be a very useful resource not only for the ethnic groups themselves but also general audiences. Ultimately, the workshops successfully trained lecturers and master's students from local communities such as Pattani Malay and Tak Bai, turning them into accomplished digital archivists with the competence to conduct all processes from collecting and managing to archiving.

It is worth noting that the International Conference on Language Diversity and Preservation of Cultural Heritage (LCP 2019) at RILCA, Mahidol University was unable to attract a large number of participants partly because language documentation and archiving is not a well-known field of study in Thailand and despite UNESCO declaring 2019 the International Year of Indigenous Languages with the objective of raising awareness of the urgent need to preserve, revitalize, and promote indigenous languages around the world. This issue has been addressed at numerous international conferences focusing on indigenous languages, including the 6th International Conference on Language Documentation & Conservation (ICLDC): Connecting Communities, Languages, & Technology, from February 28 to March 3, 2019 in Mānoa, Hawai'i;[13] International Conference on the Intersections of Language and Nature: Conservation, Documentation, and Access, from September 6 to 7, 2019 at

---

[13]https://icldc6.weebly.com/

University of Pittsburgh;[14] and Inclusion, Mobility, and Multilingual Education Conference, from September 24 to 26, 2019 in Bangkok, Thailand.[15] As a result, LCP 2019 attracted few presenters from international countries, and being the first time that Mahidol University had held such a conference on archiving, it was not well known to scholars in this particular field.

6. Summary and discussion "The Ethnic Group Digital Archive Project: Promoting the protection and preservation of language and culture diversity in Thailand" had as its main objectives: (i) to create a digital archive platform that was user-friendly, community-oriented, and suitable for local communities in Thailand and (ii) to disseminate archiving knowledge in order to promote the sustainable preservation of language and cultural diversity. Working collaboratively with ELAR, SOAS University of London, the RILCA research team was equipped with requisite skills for language documentation, data management, and archiving in order to achieve the stated objectives.

The community-oriented digital archive LangArchive-TH was implemented with priority given to the target users and ethnic community members, and for this reason, the website was designed to be simple enough to easily navigate and browse through the digital heritage resources. Comparing LangArchive-TH with other digital archives in Thailand, including SAC and Navanurak, one of the distinguishing features of its implementation was the collaboration between researchers and the community members in accordance with the participatory model, especially in the language documentation process. The native speakers, who were also language consultants, were fully involved in selecting their language and cultural heritage for recording and archiving. In other words, LangArchive-TH elicited and archived natural language data that authentically reflect their beliefs, way of life, and perspectives in order to preserve and transmit local knowledge, language, culture, and traditions to the younger generation.

Apart from community-based data, LangArchive-TH also archived academic data from scholars such as university lecturers, teachers, and master's degree students, who deposited linguistic practice data that was already transcribed in IPA or an orthography with Thai-based script. Such data can be used for knowledge transmission or educational purposes as they form the basis of Mother Tongue Based Education development.

In conclusion, the ethnic group digital archive project was a community-based research on ethnic groups in Thailand, yielding language and cultural data that would benefit not only the researchers but also the contributors and their communities. Specifically, the speaker communities, who were recognized as the prime beneficiaries of the archive, had full access and rights to their data, which allowed them to supplement, modify, and improve the digital heritage information. Moreover, the ethnic digital archive project facilitated transmission of knowledge through training

---

[14]https://www.iln2019.com/
[15]https://www.asiapacificmle.net/conference/2019

workshops and archiving manuals so as to ensure the sustainable preservation of language diversity and to encourage the community members to record their language and cultural heritages by themselves.

References

Association of Newfoundland and Labrador Archives (ANLA). 2020. (http://anla.nf.ca/)

Catawba Cultural Preservation Project. 2020. *Catawba Indian Nation archives.* (http://catawbaarchives.libraries.wsu.edu/)

Choosri, Isara. 2007. *Investigating contact-induced language change: Cases of Chung (Saoch) in Thailand and Cambodia.* Nakhon Pathom, Thailand: Mahidol University. (Doctoral dissertation.)

Department of Linguistics, University of California, Berkeley. 2020. *The California Language Archive.* (https://cla.berkeley.edu/)

Diffloth, Gérard. 1984. *The Dvaravati Old Mon language and Nyah Kur.* Bangkok: Chulalongkorn University.

Henke, Ryan E. & Andrea L. Berez-Kroeker. 2016. A brief history of archiving in language documentation, with an annotated bibliography. *Language Documentation & Conservation* 10. 411–457. ()https://scholarspace.manoa.hawaii.edu/bitstream/10125/24714/1/henke_berez.pdf)

Himmelmann, Nikolaus. 2006. Language documentation: What is it and what is it good for?. In Gippert, Jost, Nikolaus P. Himmelmann, & Ulrike Mosel (eds.), *Essentials of language documentation* (Trends in Linguistics. Studies and Monographs 178), 1–30. New York: Mouton de Gruyter.

Holton, Gary. 2012. Language archives: They're not just for linguists any more. In Seifart, Frank, Geoffrey Haig, Nikolaus P. Himmelmann, Dagmar Jung, Anna Margetts, & Paul Trilsbeek (eds.), *Potentials of language documentation: Methods, analyses, and utilization* (Language Documentation & Conservation Special Publication No. 3), 111–117. Honolulu: University of Hawai'i Press. (https://scholarspace.manoa.hawaii.edu/handle/10125/4523)

Joint Faculties of Humanities and Theology, Lunds University. 2020. *Repository and Workspace for Austroasiatic Intangible Heritage (RWAAI).* (https://projekt.ht.lu.se/rwaai)

Malone, Susan E. 2010. *Planning mother tongue-based education programs in minority language communities: Resource manual for speakers of minority languages engaged in planning and implementing mother tongue-based education programs in their own communities.* Dallas, TX: SIL International. (https://www.sil.org/system/files/reapdata/37/28/09/3728095590661524300669143094078049001 3/MLE_Program_Planning_Manual.pdf)

Malone, Susan & Dennis Malone. 2013. *Two-tract approach to reading and writing in students' first language (L1).* Dallas, TX: SIL International.

(https://www.sil.org/sites/default/files/files/2-track_method_teachers_trainers_-guide_2013.pdf)

Max-Planck-Gesellschaft. 2020. *Das archiv der Max-Planck-Gesellschaft.* (https://www.archiv-berlin.mpg.de/)

National Electronics and Computer Technology Center (NECTEC). 2020. *Navanurak.* (http://www.navanurak.in.th/)

Premsrirat, Suwilai. 2000. Development of the Chong writing system. *Journal of Languages & Cultures* 19(2). 5–18.

Premsrirat, Suwilai. 2002. Report on the Chong revitalization workshops. *The Mon-Khmer Studies Journal* 32. 175–179.

Premsrirat, Suwilai & Mirinda Burarungrot. 2018. Pattani Malay-Thai bi/multilingual education in Thailand's Deep South. In Premsrirat, Suwilai & David Hirsh (eds), *Language revitalization: Insights from Thailand*, 239–249. Bern: Peter Lang AG.

Premsrirat, Suwilai & Nattamon Rojanakul. 2015. Chong. In Jenny, Mathias & Paul Sidwell (eds.), *The handbook of Austroasiatic languages*, 601–640. Leiden: Brill.

Premsrirat, Suwilai, Siripen Ungsitipoonporn, & Isara Choosri. 2007. *The experience of language revitalization in Thailand: A case study of Chong people in Chanthaburi province.* Thailand Research Fund (TRF) District Office.

Premsrirat, Suwilai, Sujaritlak Diphadung, Ekapong Suwannaket, Isara Choosri, Sophana Srichampa, Buasuang Apinya, & Thawornpat Mayuree. 2004. *Ethnolinguistic maps of Thailand.* Bangkok: Office of the National Culture Commission.

Premsrirat, Suwilai & Uniansasmita Samoh. 2012. Planning and implementing Patani Malay in bilingual education in Southern Thailand. *Journal of the Southeast Asian Linguistics Society (JSEALS)* 5. 85–96.

Princess Maha Chakri Sirindhorn Anthropology Centre. 2020. *Database.* (https://www.sac.or.th/main/en/database)

Samoh, Uniansasmita. 2018. Identity issue through the lens of language and scripts. In Premsrirat, Suwilai & David Hirsh (eds.), *Language revitalization: Insights from Thailand*, 239–249. Bern: Peter Lang AG.

SOAS University of London. 2020. *The Endangered Language Archive (ELAR).* (https://elar.soas.ac.uk/)

Thawornpat, Mayuree. 2018. Lavue revitalization project and networks. In Premsrirat, Suwilai & David Hirsh (eds.), *Language revitalization: Insights from Thailand*, 143–153). Bern: Peter Lang AG.

Ungsitipoonporn, Siripen. 2011. Development of Northern Khmer primer using Thai alphabet: Opportunities and challenges. In Srichampa, Sophana, Paul Sidwell, & Kenneth Gregerson (eds.), *Austroasiatic studies: Papers from ICAAL 4* (The Mon-Khmer Studies Journal Special Issue No. 3), 134–151.

Ungsitipoonporn, Siripen. 2018. The impact of the Nyah Kur Language revitalization. In Premsrirat, Suwilai & David Hirsh (eds.), *Language revitalization: Insights from Thailand*, 81–103. Bern: Peter Lang AG.

Ungsitipoonporn, Siripen (ed.). 2020. *The digital archive for the languages and cultures of the ethnic groups of Thailand.* Bangkok: Thailand Science Research

and Innovation (TSRI). (https://langarchive-th.org/th/system/files/atoms/file/Archiving_book.pdf)

Ungsitipoonporn, Siripen & Kumaree Laparporn. 2019. Comparison of approaches for language revitalization of Northern Khmer in Thailand. *International Journal of Society, Culture & Language* 7(1). 52–66. (http://www.ijscl.net/article_34782.html)

Siripen Ungsitipoonporn
siripen.ung@mahidol.edu
orcid.org/0000-0002-7685-8498

Buachut Watyam
buachut.bcw@gmail.com

Vera Ferreira
vf4@soas.ac.uk

Mandana Seyfeddinipur
ms123@soas.ac.uk