

# PARADISEC 2025 ANNUAL REPORT



Apu Kalsarap Nemaf and Aiti Limaas Kalsarap reading a dictionary of their language. Erakor village, Vanuatu, 2001.

The Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC) continues to accession new materials in a number of languages. This report outlines activity in 2025 as we worked on new collections, added to existing collections, produced podcasts, and improved our catalog and ingestion system. We are exploring developments in playback of media together with transcripts, and the use of automatic speech recognition and machine translation that will increase access to information in the collection.

# Collection Metrics

TOTAL SIZE 295 TB

	12/2023	12/2024	12/2025
COLLECTIONS	680	749	811
ITEMS	37,427	41,967	45,961
FILES	427,269	437,916	470,561
VIDEO FILES (VID-MP4)	14,256	15,742	18,144
HOURS OF AUDIO	16,500	17,600	19,544

## Awards

- ▶ Nick Thieberger, Amanda Harris and their co-authors received an ASA Mander Jones Award Commendation for “The New Protectionism: Risk Aversion and Access to Indigenous Heritage Records”. Michael Aird, Clint Bracknell, Jason Gibson, Amanda Harris, Marcia Langton, Gaye Sculthorpe, Jane Simpson and Nicholas Thieberger. Archives & Manuscripts Vol. 51(2) 2023, pp 23-42  
[10.37683/asa.v51.10971](https://doi.org/10.37683/asa.v51.10971)
- ▶ We won the Open Scholarship Award sponsored by the Canadian Social Knowledge Institute. Nick Thieberger accepted the award at the Canadian Australian Partnership for Open Scholarship (CAPOS) 7th annual gathering in Canberra in December. The award notes that recipients demonstrate exemplary open scholarship via research, projects, or initiatives. These awards are intended to acknowledge and celebrate exemplary open scholarship, nominated via an open process.

## News Articles

- ▶ Nick Thieberger visited the Vanuatu Cultural Centre, returning newly digitised tapes and picking up more for future digitisation. Two articles discuss Nick's ongoing work with the Centre:  
<https://vbtv.vu/vanuatu-cultural-heritage-tapes-being-digitized-for-online-access/>  
[https://www.dailypost.vu/features/bringing-voices-home---the-duo-digitising-vanuatu-s-cultural-history/article\\_162ebb5e-834d-4562-a0ab-09c2eb7da4d1.html](https://www.dailypost.vu/features/bringing-voices-home---the-duo-digitising-vanuatu-s-cultural-history/article_162ebb5e-834d-4562-a0ab-09c2eb7da4d1.html)
- ▶ ABC Pacific Beat broadcast a radio story about Pacific communities reconnected with traditional songs thanks to PARADISEC's digitisation efforts, featuring Amanda Harris and Steven Gagau:  
<https://www.abc.net.au/pacific/programs/pacificbeat/pacific-music-recordings/105842576>  
A TV segment was also included on ABC The Pacific on 2 October 2025.

## Blog

- ▶ We published seven blog posts since the last report:
  - [The new PARADISEC catalog viewer](#)
  - [Speech recognition on your laptop](#)
  - [Large language models for small languages](#)
  - [Splicing, cracking, rehousing: Preparing cassettes for digitisation](#)
  - [Redeveloped uploading for PARADISEC](#)
  - [The Insider Archivist: Collecting Music Recordings from East New Britain \(ENB\)](#)
  - [The Tape Restorator](#)

## Podcast

- ▶ Our podcast series Toksave continued in 2025 with several episodes produced ready for release early 2026:
  - Episode 16: The music of Malaita, based on the IF01 collection, interviews with Ian Fraser and Malaitan community member Mary Sattin
  - Episode 17: Bamboo Music of the Phillipines, a special episode produced by USyd PhD candidate Starr Abelardo along with audio engineer Michael Manzini. Interviews with Waway Saway and Rhyam Clapham (Dobby)
  - Episode 18: Opening the archive: Indigenous perspectives on archival engagement - A special episode based on audio recorded at panel with Steven Gagau, Lauren Booker, Johnny Obed and Evelyn Quispe, moderated by Amanda Harris
  - Episode 19: Issues for Pacific Archives: A special episode recorded at the PARBICA 21 conference.

## Funding

- ▶ The Language Data Commons of Australia is funding part of our operations until 2028
- ▶ Julia Miller secured approx AUD \$98,000 from the Modern Endangered Archives Program (MEAP) which will be used to inventory, digitise, and archive the David Walsh collection, including audio reels, filed notebooks, unpublished manuscripts, lexical cards, and photos collected in Vanuatu, Papua New Guinea, Cook Islands, and French Polynesia. (2 years of RA funding for ~5 new casual staff members).
- ▶ The Sydney team were awarded a University of Sydney internal DVCR CAPEX Digital Allowance Grant, for \$33,531 to refresh our field recording equipment and support testing of the updated raspberry pi RO-Crate file delivery.
- ▶ PDSC (Inc) funds: Income from donations was around \$10,000. The current balance is \$18,765. From this we paid:
  - Staff attendance at conferences
  - Printing of 300 new teatowels as promotional material
  - New postcards as promotional material
  - PARBICA conference registration
  - Greenfleet carbon offset
  - IASA subscription
  - We donated \$1000 to PARBICA to sponsor Pacific attendees for travel costs to the 2025 PARBICA conference in New Zealand.

## Collections added/significant new content added to established collections

3PAC1 3PAC2 3PAC3 - transcribed Tok Pisin, Solomons Pijin and Bislama

AB2 - Dani, West Papua

AC1 - Arthur Capell - ELAN summaries

AJ1 - Mekeo, Papua New Guinea

AS4 - Kamberau, Malay, Indonesia

AV1 - Javanese, Bali, Indonesia

BA2 - Huli, Papua New Guinea

BB2 - Pitta Pitta, Australia - ELAN summaries

BL1 - Naga, Wancho, India

BSC1 - Bislama Spoken Corpus

BJM02 - Kaur, Nasal, Indonesia

BK1 - Barb Kelly, Sherpa, Nepal

CCLD1 - Milang, India

CCLD2 - Dakpakha, Tshangla, Bhutan

CCLD3 - Hindi, multiple languages, India

CCLD4 - Luoba, Yidu, India

CCLD5 - Darang Deng, India

CCLD6 - Yobin/Lisu, India

CCLD7 - Hruso, India

CCLD8 - Apatani, India

COCAS - Assamese, India

DB2 - Nasu, Sanyie, China

DD1? - Papuan languages

DO01 - Siwai, Papua New Guinea

DO02 - multiple languages, Papua New Guinea

DP1 - Balik, Indonesia

DW1 - Sekar, Indonesia

ECM - Pulabu, Tok Pisin, Papua New Guinea

EMG1 - multiple languages, Papua New Guinea

EQ1 - Warlpiri, Australia

ER2 - Bislama, Ambrym, Vanuatu

ER3 - Bislama, Ambrym, Vanuatu

ES1 - Suau, Papua New Guinea

EV2 - Onim, Papua

SDM23 - Naga, Tase, India

SG2 - Steven Gagau, In progress

SINM - Solomon Islands National Museum

SMR1 - Devayan, Indonesia

SN1 - multiple languages, New Zealand, Cook Islands

TD4 - Naga, Wancho, India

FD1 - Warlpiri, Australia

HK1 - Wasembo, Tok Pisin, Papua New Guinea

ICTMD48 - performances at the ICTMD conference 2025

IG3? - typed notebook, Magati Ke, Australia

JAG2 - Anmatyerr, Australia

JAP1 - Geser-Gorom, Indonesia

JM5 - Marri Tjevin, Australia

JM7 - Mum, Papua New Guinea - no data yet

GL1 - Huli, Papua New Guinea

KD2 - Wagiman, Australia

KI01 - Sibe, Tok Pisin, Papua New Guinea

KJ1 - Djambarrpuyuṅu, Australia

KK1, KK2, KK3 - Kachin, Myanmar

LL1 - Bislama, Vanuatu

MB2 - Lakoya, South Sudan

MD8 - Dusner, Indonesia

MD9 - Biak, Indonesia

MLC2 - multiple languages, Australia

MO01 - Siwai, Papua New Guinea

MO02 - Siwai, Papua New Guinea

MO03 - Naasioi, Tok Pisin, Papua New Guinea

ILG01 - Marrgu, Ilgar, Australia

MR1 - multiple languages, Indonesia, Papua New Guinea, Solomon Islands

OCSEAN2 - multiple languages, Indonesia, Papua New Guinea

PAMBU - multiple languages, multiple countries

PL2 - Siberia

PM1 - Arguni, Indonesia

RR2 - Kawi Old Javanese, Bali, Indonesia

SA1 - multiple languages, Philippines

SB03 - Kwamera, Vanuatu

SDM16 - Rebab, Serunai, Malaysia

SDM19 - Naga, Tase, India

TU1 - Gyegu Tibetan, China - no data yet

WM2 - Wantok Musik Foundation (WMF) - Music Label Recordings of Oceania

VKS4 - multiple languages, Vanuatu

WZ1 - no data yet

YDK1 - Iyaq Nam, Papua New Guinea

## Large collections returned

- ▶ Nick Thieberger returned SINM tapes to the Solomon Islands National Museums, and had talks about setting up a language institute and archive
- ▶ Nick Thieberger returned VKS tapes to the Vanuatu Cultural Centre and brought more back for digitisation - see news articles above

## Conferences

- ▶ We are co-organisers of the Language Documentation and Archiving (LD&A) conference, to be held in Berlin in September, 2026 <https://langdoc.org/lda26/>

## Equipment

- ▶ Sam King delivered the cassette lubrication machine to Sydney and trained staff to use it
- ▶ The Studer reel-to-reel machine in the Sydney office has been repaired (replacement capstan motor was purchased from France after much searching for these rare parts).
- ▶ Both cassette machines in the Sydney office have also been serviced.
- ▶ Julia Miller has created PARADISEC branded 3D-printed reel-to-reel tape clips



## Processes

- ▶ We have launched a new [PARADISEC catalog](#) interface (February 2026) using the platform Oni. This work has been funded via the LIEF grant which ended in 2025 as well as LDACA funding.
- ▶ Nick Thieberger has been working on an Automatic Speech Recognition project for Bislama with a student in the computer science department at the University of Melbourne (it is being run on the closed Uni Melb system). 22 hours of Bislama has now been transcribed using ASR, and an auto-translation has also been done. Everything has been manually checked and the accuracy of the ASR is really good. A second ASR transcription project is underway for Nafsan. Results are more variable due to the quality of the recordings. The ability to now run these ASR projects has implications for audio in PARADISEC which has little metadata.
- ▶ We have developed a new video processing pipeline: <https://github.com/paradisec-archive/paragest>
- ▶ We have added a new privacy notice for EU Users on website: <https://www.paradisec.org.au/european-union-general-data-protection-regulation-gdpr/>
- ▶ We are setting up new email list which will be rolled out with the new Oni instance in early 2026
- ▶ Work is underway to allow a user to select items from the catalog and to export them in a format that can be put onto a Raspberry PI, replicating what we did in the past with 'Dataloader', but now conforming to the LDACA RO-Crate standard, so allowing others to use the coding in their own collections.

## Publications

Jodie Kell completed her PhD thesis, which links to the JK1 PARADISEC collection. Other PhD students are doing the same thing:

Jodie Kell (2025). The Ripple Effect of the First All-Women's Band from West Arnhem Land, PhD Thesis, Sydney Conservatorium of Music, University of Sydney.  
<https://hdl.handle.net/2123/34247>

Amanda Harris, Deveni Temu, Steven Gagau and Jodie Kell. (2025). Archived sound and creative engagements with Papua New Guinean cultural heritage in Australia. *History Australia*, 1–21.  
<https://doi.org/10.1080/14490854.2025.2559634>

Nick Thieberger (2025). Records of Pacific Languages: Where Are They and Who Can See Them? Department of Pacific Affairs In Brief Series.  
<https://doi.org/10.25911/4Q9D-VM92>

Nick Thieberger wrote an Arena article about the importance of preserving humanities research data and the establishment and ongoing development of PARADISEC  
<https://arena.org.au/public-good-curation-of-humanities-research/>

## Presentations

Amanda Harris, Nick Thieberger and Steven Gagau, 'Archiving and preservation in the age of digital sharing', Association for Social Anthropology in Oceania, Fiji and online, 13 February, 2025.

Nick Thieberger was invited to present on Papuan languages and PARADISEC at the Out of Papua conference at the Tokyo University of Foreign Studies, 17 February, 2025.

Nick Thieberger was invited to present on PARADISEC and digital language archiving at Kyoto University, 20th February, 2025.

Nick Thieberger. The Importance of Digital Language and Music Archives for Language and Cultural Continuity: an Example from the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC), 2nd International Conference on Language Technologies for All (LT4All 2025), UNESCO Headquarters, Paris, 24 - 26 February, 2025.

Nick Thieberger, 'Improving access to and sustainability of the Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC)' Digital Preservation Masterclass: A Celebration of Good Practice, Innovation and Collaboration, online, 12th March, 2025

<https://www.dpconline.org/doclink/dpmasterclass-2025-paradisec/>

Mandana Seyfeddinipur, Anasuya Sengupta, Serge Sagna, Nick Thieberger and Valentina Vapnarsky. 2025. 'Taking a global perspective: Who can actually discover and access born digital collections'. Panel discussion at the conference [Born-Digital Collections, Archives and Memory](#), online and University of London, 4th April 2025.

Nick Thieberger. 'Taking Language Archiving Seriously'. Talk given in the Department of Linguistics at Cologne University, April 17th, 2025.

Nick Thieberger. Georeferencing for access and discovery of language data. 1st meeting of the Establishing Spatiotemporal Linguistics: A New Approach to the Interdisciplinary Study of Language History and Human Movement. April 27th, 2025, National Museum of Ethnology, Osaka.

Steven Gagau. PARADISEC stand at Australian Association for Pacific Studies Conference, University of Sydney, 3 - 6 June, 2025.

Nick Thieberger, Danielle Barth, Bethwyn Evans, Nicholas Evans, Kirsty Gillespie, Raphael Merx, Sam Passmore & Lila San Roque. Building and using corpora for Pacific languages: examples from the Pacific Creoles Project. 13th Conference on Oceanic Linguistics (COOL13), Canberra, 23 - 27 June, 2025.

Jodie Kell and Steven Gagau. Opening the Archive: Access, Engagement and Innovation Exhibition, Sydney Conservatorium of Music Library, Sydney, August to November 2025.

Steven Gagau, Dr Lauren Booker, Johnny Obed, Evelyn Quispe and Amanda Harris. Opening the archive: Indigenous perspectives on archival engagement, Panel discussion, Sydney Conservatorium of Music Library, 8th October, 2025.  
<https://www.idaca.edu.au/news/posts/paradisec-panel-oct2025/>

Steven Gagau attended the ICDH/UNESCO International Workshop on the Memory of the World Programme for Pacific Islands, Sydney, 28 - 30th October, 2025.

Jodie Kell & Steven Gagau. PARADISEC Podcast Series, TOK-SAVE: Culture Talks. Lightning Talk/Poster presentation, 21st International Conference on Digital Preservation, Wellington, New Zealand, 3 - 7 November, 2025.  
<https://drive.google.com/file/d/1nN3lOserFLWvLk5fMoFb-Vu1xde5leUFy/view?usp=sharing>

Jodie Kell, Steven Gagau and Julia Miller. Collaboration, Education, Engagement: Digital Archiving in Action Keynote Address at PARBICA 21 – Demystifying Digital, Wellington, New Zealand, 7 - 8 November, 2025.

Jodie Kell, Steven Gagau and Julia Miller. From Start to Finish: A Workflow for Digital Archiving, PARADISEC Workshop, PARBICA 21 – Demystifying Digital, Wellington, New Zealand, 7 - 8 November, 2025.

Steven Gagau. "Insider Archivist": Initiating a community archiving project, Australian Society of Archives Conference, 10 - 13 November, 2025.

Peter Sefton, Nick Thieberger and Michael Falk. RO Crate for a data commons, Digital Humanities Australasia Conference, ANU, Canberra, 3 - 5 December, 2025.

Nick Thieberger. "The Written and the Recorded: AI in the workflow from analogue materials to textual corpora", Canadian Australian Partnership for Open Scholarship (CAPOS) Conference, "Re-Defining Open Social Scholarship in an Age of Generative 'Intelligence'", ANU, Canberra, 2 - 3 December, 2025.

## Teaching, Training, Workshops

- ▶ Julia Miller gave a guest lecture for PASI1012, introducing Pacific collections of ANU to undergrad students.
- ▶ Julia Miller gave a guest lecture for LING9001 (Topics in Linguistics) on collaborative approaches to archiving that support cultural continuity and Indigenous authority.
- ▶ Julia Miller trained up two new RAs in the Canberra office
- ▶ The Sydney office hosted visits from:
  - ▶ University of Sydney musicology and ethnomusicology classes  
ICDH International Workshop on the Memory of the World for Pacific Islands (at which Steven Gagau was also a delegate on behalf of PARADISEC)
  - ▶ LDaCA staff member Bridey Lea, who had a professional development day at the office and wrote a facebook post about it <https://www.facebook.com/reel/1168684168041810>
- ▶ Steven Gagau attended the ARDC 2025 HASS and Indigenous Research Data Commons Summer School, 3 - 6 February, 2025, and was profiled in this article.  
<https://ardc.edu.au/article/recap-of-the-2025-hass-and-indigenous-research-data-commons-summer-school/>
- ▶ Steven Gagau attended the Colonial Heritage in the Pacific workshop in Sydney, 19 - 20 June, 2025.
- ▶ Jodie Kell gave a guest lecture for the Intercultural Music Seminar PERF5005 Principal Study 1 at the Sydney Conservatorium of Music

## Service

- ▶ Nick Thieberger is the President of the Digital Endangered Languages and Musics Archives Network (DELANAN). He is currently leading their work to revamp OLAC <https://www.delaman.org/resources/olac-revamp/>
- ▶ Julia Miller continues to serve on 2 committees at IASA (the International Association of Audiovisual Archives):
  - Technical Committee
    - ▶ Also member of the subcommittee dedicated to organising and creating the IASA-TC07 TC Report: Guidelines for the Preservation of Born-digital Video
  - Diversity, Equity, and Inclusion Committee
- ▶ Steven Gagau was elected to the PARBICA Bureau
- ▶ Jodie Kell volunteered with the PARBICA Bureau
- ▶ Steven Gagau serves as a board member of Wantok Musik Foundation





If you wish to make an impact by supporting PARADISEC's projects and operations, please donate to PARADISEC (Inc). PARADISEC (Inc) is a deductible gift recipient which means that any donations over \$2 can be claimed as a tax deduction. You can also donate online here:  
<https://paradisec.org.au/support-us>.



THE UNIVERSITY OF  
SYDNEY



Australian  
National  
University



THE UNIVERSITY OF  
MELBOURNE



paradisec.org.au