Access to recordings in the languages of the Pacific

Nick Thieberger
Director of PARADISEC
School of Languages and Linguistics
The University of Melbourne

PARADISEC acknowledges and pays respect to the original owners of the land we work on, the Gadigal people of the Eora Nation, the Ngunawal and the Woiwurrung
A quarter of the world’s languages are in the Pacific
Analog records not being looked after
How to find that recording made with your grandmother in the 1960s?
Legacy collections with little metadata
“Today audiovisual archives associations estimate the time window still open for the transfer of dedicated analogue and digital carriers into digital repositories to be not more than just 20 years.”

“Preservation alone is not enough: an ethnographic legacy packed in shoe boxes and tea chests across the six inhabited continents is no legacy at all.”

Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC)

Collaborative digital research resource set up by University of Sydney, University of Melbourne & Australian National University, 2003 with the aim of preserving ethnographic (language, oral history, music) recordings made since the 1950s.

funding from Australian Research Council
Why PARADISEC?

No agency took responsibility for audio recordings made in the Pacific

An archive built by researchers, not archivists
PARADISEC

1,177 languages
449 collections
20,911 items
238,928 files
8,778 hours (audio)
775 hours (video)
48.3 TB of files

23.4 TB of video files in archive
13.3 TB of audio files in archive
1.16 TB of image files in archive
1.16 GB of text files in archive
Solomon Islands Museum: > 600 tapes, most include oral tradition in local languages. 2014 – Endangered Archives Programme funding to digitise 200 tapes
Returning 100 tapes to the Vanuatu Kaljoral Senta

Ongoing work 2017-2018 to digitise a further 100 tapes
API, metadata feeds picked up by Research Data Australia, OAI services, Open Language Archives Community (OLAC), etc.
API, metadata feeds picked up by the Virtual Language Observatory (CLARIN)
OLAC resources in and about the South E fate language

ISO 639-3: erk

The combined catalog of all OLAC participants contains the following resources that are relevant to this language:

- **Primary texts**
  - Lexical resources
  - Language descriptions
  - Other resources about the language
  - Other resources in the language

Other known names and dialect names: South E fate, Erakor, Pate, Southern E fate

Use faceted search to explore resources for South E fate language.

**Primary texts**

1. Open Language Archives Community (OLAC) [Open Language Archives Community (OLAC)]
   - http://www.language-archives.org/language/erk
   - Using ISO-639-3 codes for languages

**Lexical resources**

1. Open Language Archives Community (OLAC) [Open Language Archives Community (OLAC)]
   - Using ISO-639-3 codes for languages

**Language descriptions**

1. Open Language Archives Community (OLAC) [Open Language Archives Community (OLAC)]
   - Using ISO-639-3 codes for languages

**Other resources about the language**

1. Open Language Archives Community (OLAC) [Open Language Archives Community (OLAC)]
   - Using ISO-639-3 codes for languages
go.coedl.net/OLACviz
How much is in the archives for languages of Vanuatu?
What else is out there?

- Items held in personal collections can’t be located:
  - speakers who recorded their families
  - missionaries
  - patrol officers
  - academics
- Enter these into our catalog to make them findable

- Using ISO-639-3 and OLAC metadata
The New Testament ... translated into the language of Efate, New Hebrides. (Tusi tab fau ni lesu Kristo, etc.) [Translated by D. Macdonald and J. W. Mackenzie.]
Melbourne : B. & F.B.S., 1889.
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform Title</td>
<td>Bible. N.T. Efate.</td>
</tr>
<tr>
<td>Also Titled</td>
<td>New Testament of our Lord and Saviour Jesus Christ translated into the language of Efate, New Hebrides</td>
</tr>
<tr>
<td>Other Authors</td>
<td>British and Foreign Bible Society. Victoria Auxiliary</td>
</tr>
<tr>
<td>Published</td>
<td>Melbourne : Printed for the British and Foreign Bible Society by the Victoria Auxiliary, 1889.</td>
</tr>
<tr>
<td>Physical Description</td>
<td>394 p.; 19 cm.</td>
</tr>
<tr>
<td>Subjects</td>
<td>Efate language -- Texts.</td>
</tr>
<tr>
<td>Language</td>
<td>Austronesian (Other)</td>
</tr>
<tr>
<td>Dewey Number</td>
<td>220.5994</td>
</tr>
<tr>
<td>Libraries Australia ID</td>
<td>24155203</td>
</tr>
<tr>
<td>Contributed by</td>
<td>Libraries Australia</td>
</tr>
<tr>
<td>Beschriftungsort</td>
<td>a) Plattenlabel, Mitte; b) in die Platte geresst, mittig</td>
</tr>
<tr>
<td>-----------------------</td>
<td>--------------------------------------------------------</td>
</tr>
<tr>
<td>Format</td>
<td>27 cm</td>
</tr>
<tr>
<td>Hauptsprecher / Sänger</td>
<td>Williams, Salomon</td>
</tr>
<tr>
<td>Muttersprache</td>
<td>Tahitisch (Polynesisch)</td>
</tr>
<tr>
<td>Aufnahmedauer</td>
<td>min.</td>
</tr>
<tr>
<td>Land (historisch)</td>
<td>Australien</td>
</tr>
<tr>
<td>Land (heute)</td>
<td>Australien (Land)</td>
</tr>
</tbody>
</table>

Tahitisch (Australien), Gespräch - PK 900/2
Existing resources = low-hanging fruit

Noa‘ia ‘e Mauri

Welcome to a website for the Rotuman people

Rotuma is a volcanic island of approximately 43 sq. kilometers, located at 12 degrees south latitude and 177 degrees east longitude, approximately 465 kilometers north of Fiji. Although the island has been politically part of Fiji since 1881, Rotuman culture more closely resembles that of the Polynesian islands to the east, most noticeably Tonga, Samoa, Futuna and Uvea. Because of their Polynesian appearance and distinctive language, Rotumans now constitute a recognizable minority group within the Republic of Fiji.
The NATIVES of ROTUMA. By J. STANLEY GARDINER, B.A.,
Gonville and Caius College, Cambridge. (Communicated
by Professor ALEXANDER MACALISTER, M.A., F.R.S.)

[J. Stanley Gardiner, a naturalist, visited Rotuma for 3 1/2 months in 1896. He was a keen observer and conducted extensive interviews on a range of topics. His account, published in 1898 in the Journal of the Royal Anthropological Institute 27:396-435, 457-524, is the most comprehensive ethnography of Rotuma published in the nineteenth century. It is an indispensable source for studies of Rotuman culture and history.]

Note: Any citations drawing on this material should acknowledge the website address <http://www2.hawaii.edu/oceanic/rotuma/os/hanua.html>, the date of downloading, and full citation of the original publication.

CONTENTS.

Introsuction XIV. The Sou and his Officers.
I. Discovery and Historical, XV. Religion.
II. Early Canoe Voyagers, XVI. Warfare.
III. Physical and Moral Characteristics, XVII. Cannibalism.
IV. Dress, Ornaments, and Tattooing, XVIII. Marriage.
V. Mats, Baskets, Thatch, and Sinnet, XIX. Tenure of Land.
VI. Planting, Food, and Cookery, XX. Sports, Games, and Toys.
VII. The Kava Ceremonial, XXI. Singing and Dancing.
VIII. Hooks, Nets, and Fishing, XXII. Medicine and Surgery.
IX. Districts and their Government, XXIII. Decrease of the Native Population.
The NATIVES OF ROTUMA. By J. STANLEY GARDINER, B.A.,

[J. Stanley Gardiner, a naturalist, visited Rotuma for 3 1/2 months in 1896. He was a keen observer and conducted extensive interviews on a range of topics. His account, published in 1898 in the Journal of the Royal Anthropological Institute 27:396-435, 457-524, is the most comprehensive ethnography of Rotuma published in the nineteenth century. It is an indispensable source for studies of Rotuma culture and history.]
Olac resources in and about the Rotuman language

ISO 639-3: rtm

The combined catalog of all OLAC participants contains the following resources that are relevant to this language:

- Primary texts
- Lexical resources
- Language descriptions
- Other resources about the language
- Other resources in the language

Other known names and dialect names: Rotuna, Rotuman

Use faceted search to explore resources for Rotuman language.

Primary texts


Lexical resources

Making Pacific Languages Discoverable: Blending Library and Language Metadata Standards at the University of Hawai‘i at Mānoa Library Pacific Collection

Eleanor Kleiber, Andrea L. Berez-Kroeker, Mike Chopey, Ryan Shelby, & Danielle Yarbrough (U Hawai‘i at Mānoa)

Rationale

The UHMI Library Pacific Collection is the world’s premier collection of languages for research on Pacific Languages. The collection contains more than 20,000 items on the nearly 1,100 languages of Micronesia, Polynesia, and Melanesia.

The language holdings in the Collection include dictionaries, grammars, and folk materials, as well as extensive material written in Pacific Languages, such as religious texts, government documents, children’s books, and public health materials. Many of these materials are unique and rare and may not be found in other libraries in the world. The Collection represents almost all the Pacific Islands languages for which materials have been formally or informally published, making it an invaluable resource for the study of Pacific Islander languages.

Before this project began in 2014, we didn’t know exactly how many and which languages were represented in the Collection. After two years, the first result is that there are all of the following record titles in the holdings of the Pacific Collection:

- “A language description for the X...”
- “A language description for the Y...”
- “A language description for the Z...”

The second result is that the collection contains approximately 100,000 catalog records in the cataloging standards accepted by library science and linguistics science.

Project Partners

The Pacific Collection at the UHMI Hamilton Library is the pioneer for the type of language and cultural materials for research on Pacific Languages and Languages. In total, the Pacific Collection provides a rich mix of materials across the Pacific Islands for which materials have been formally or informally published, making it an invaluable resource for the study of Pacific Languages. While they have an extensive cataloging effort each Pacific Languages region, there are no comprehensive cataloging standards accepted by library science and linguistics science.

Mapping Library and Language Description Standards

Librarians describe languages in their catalog records in three ways:

- In the Notes field (E0) (3/22), using notes.
- In the bibliographic description (E10), using notes.
- In the container title (E11), using notes.

Examples of problems we’ve encountered and fixed

Cataloging

- During the cataloging process, we encountered several issues with the language metadata, such as:
  - Incorrect classification of the language
  - Missing or incorrect language codes
  - Lack of language descriptions

For example, the book titled “The Language of...” has a language code of “en”, which is incorrect. The correct language code should be “kik”. We fixed this by changing the language code to “kik”.

(c) 2001-2003 Copyright held by //////////////////////////////////////////////////////////////////////

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage or retrieval system, without permission in writing from the publisher.

References


This book is in Japanese...
Looking for analog tape collections

Digital Languages and Musics Archives Network
https://rebrand.ly/lostandfound
Lost and found – locating collections

Don Kulick’s collection of recordings: Tayap (PNG)
http://catalog.paradisec.org.au/collections/DK1

Zygmunt Frajzyngier: North Africa
http://catalog.paradisec.org.au/collections/ZF1

Wolfgang Sperlich: Namakira (Vanuatu)
http://catalog.paradisec.org.au/collections/WS1

Lamont Lindstrom: Kwamera (Vanuatu)
http://catalog.paradisec.org.au/collections/LL1
Lost and Found

- John Z’graggen (196 cassettes/r2r tapes in Basel Museum)
  - Carried to Nijmegen (Netherlands) and digitised at the MPI

- Stuart Kirsch (8 cassettes in Michigan, USA)
  - Language: Yonggom (PNG)
  - Sent to AILLA Texas, USA) for digitising

- Hans Fischer (61 r2r tapes in Switzerland)
  - Languages: Banir, Wotut, Jetuje (PNG)
  - Digitised at ELAR (London, UK)

All then accessioned at PARADISEC
Enriching metadata

How to get records back to their source?

- sending hard disks to cultural centres
- making our catalog maximally accessible and findable
- providing low resolution copies of archival files for delivery
- problem of providing context for a set of files
PARADISEC collection

20,911 items
238,977 files

Delivery to source community

Item 1 (20 files)
Item 2 (2 files)
Item 3 (4 files)
Item 4 (7 files)

Problem: no catalog accompanies the files
PARADISEC catalog writes xml files

<item>
  <identifier>N1-006</identifier>
  <collectionId>154</collectionId>
  <private>false</private>
  <title>Recordings in South Efate</title>
  <description>Stories of Langatalof, and Kapatol.</description>
  <creationDate>1997-11-11</creationDate>
  <originationDate>1997-11-11</originationDate>
  <creationNarrative>1997-11-19, 1997-11-11</creationNarrative>
  <accessLink>http://catalog.paradisec.org.au/coll/154</accessLink>
  <creator>Nick Thieberger</creator>
  <country>Vanuatu - VU</country>
  <language>Bi-Slana - bi</language>
  <subjectLanguages>
    <language>Bi-Slana - bi</language>
    <language>Si-Efate, South - erk</language>
  </subjectLanguages>
  <contentLanguages>
    <language>Bi-Slana - bi</language>
    <language>Si-Efate, South - erk</language>
  </contentLanguages>
  <dialect>
    <region>Emir-er village</region>
    <geographicLocation>
      <southLimit>17.8235</southLimit>
      <northLimit>17.7235</northLimit>
      <eastLimit>168.317</eastLimit>
      <westLimit>168.217</westLimit>
    </geographicLocation>
  </dialect>
</item>

<agent role="depositor">
  Nick Thieberger
</agent>

<agent role="recorder">
  Nick Thieberger
</agent>

<agent role="speaker">
  Kalsarap Namaf
</agent>

<agent role="speaker">
  Jokopeth
</agent>

<citation>Nick Thieberger (depositor, recorder), Kalsarap Namaf (speaker), Jokopeth (speaker)</citation>

<files numFiles="2">
  <file>
    <name>N1-006-006A.mp3</name>
    <mimeType>audio/mpeg</mimeType>
    <fileSize>30 MB</fileSize>
    <duration>00:32:43.109</duration>
    <bitrate>125 KB/s</bitrate>
    <sampleRate>44100 Hz</sampleRate>
    <channels>Stereo</channels>
    <frameRate/>
  </file>
  <file>
    <name>N1-006-006A.wav</name>
    <mimeType>audio/x-wav</mimeType>
    <fileSize>331 MB</fileSize>
    <duration>00:32:47.569</duration>
    <bitrate>1.35 MB/s</bitrate>
    <sampleRate>44100 Hz</sampleRate>
    <channels>Stereo</channels>
    <frameRate/>
  </file>
</files>

Many of these transcripts have been interlinearised in the version of the text in the file. The first file is the digitised recording, and the second file is the original analogue cassette.

<archiveInfo>
  <bornDigital>false</bornDigital>
  <bornDigital>true</bornDigital>
  <dateCreated>2012-09-27 20:00:00</dateCreated>
  <recordCreated>2012-09-27 20:00:00</recordCreated>
  <dateDigitised>2001-01-01 00:00:00</dateDigitised>
  <digitisationNotes/>
  <metadataImportedOn/>
  <metadataExportedOn/>
  <trackingID given to the Vanuatu Cultural Centre trackID/>
</archiveInfo>

<adminInfo>
  <adminAccess>
    <adminAccess/>
  </adminAccess>
  <dataAccessConditionsOpen (subject to agreeing to PDSC acc.)/dataAccessConditionsOpen (subject to agreeing to PDSC acc.)>
  <dataAccessNarrative/>
  <adminComment/>
</adminInfo>

<userComments numComments="0"/>
</userComments>
</collection>

<agents>
  <agent role="depositor">
    Nick Thieberger
  </agent>
  <agent role="recorder">
    Nick Thieberger
  </agent>
  <agent role="speaker">
    Kalsarap Namaf
  </agent>
  <agent role="speaker">
    Jokopeth
  </agent>
  <agent role="depositor">
    Nick Thieberger
  </agent>
  <agent role="recorder">
    Nick Thieberger
  </agent>
  <agent role="speaker">
    Kalsarap Namaf
  </agent>
  <agent role="speaker">
    Jokopeth
  </agent>
</agents>

<files numFiles="2">
  <file>
    <name>N1-006-006A.mp3</name>
    <mimeType>audio/mpeg</mimeType>
    <fileSize>30 MB</fileSize>
    <duration>00:32:43.109</duration>
    <bitrate>125 KB/s</bitrate>
    <sampleRate>44100 Hz</sampleRate>
    <channels>Stereo</channels>
    <frameRate/>
  </file>
  <file>
    <name>N1-006-006A.wav</name>
    <mimeType>audio/x-wav</mimeType>
    <fileSize>331 MB</fileSize>
    <duration>00:32:47.569</duration>
    <bitrate>1.35 MB/s</bitrate>
    <sampleRate>44100 Hz</sampleRate>
    <channels>Stereo</channels>
    <frameRate/>
  </file>
</files>
LibraryBox

- Delivery of archival files avoiding bandwidth cost
- LibraryBox is a low powered webserver
  - wifi transmitter
  - USB drive stores the data

http://librarybox.us/
Elisa Alphonse in Erakor village

Kalonuk Albert & Gino Joseph in Erakor village
Glossopticon: https://www.youtube.com/watch?v=X2iimqq7LS8

https://glossopticon.com/vr.html
PARADISEC – unlocking heritage records

Researchers can make recorded materials available more easily than in the past.

PARADISEC
- provides a platform for finding and accessing these records from anywhere in the world (decolonising?)
- actively seeking analog tape collections and funding to digitise them
- building a self-describing collection that can outlive the current catalog

Please let us know about collections of audio or film that need to be preserved
http://catalog.paradisec.org.au

Australian National Data Service

Australian Research Council – LIEF grants LE110100142, LE0560711, LE0453247
ARC DP0450342, DP0984419, & FT140100214

ARC Centre of Excellence for The Dynamics of Language