Language documentation and ethnography

Norwegian University of Science and Technology (NTNU)
Trondheim, September 11th, 2013

Nick Thieberger
University of Melbourne
A little history
Language documentation – what is it?
Problems of grammatical descriptions of languages
Reaction to structuralist theory
A postcolonial linguistics?
Details of documentation – methods
Archiving as a central part of documentation
If I could, by a good phonographic record, counterfeit the living voice of Tokulubakiki: [...] I should certainly be better able to translate the text in the sense of imparting to it its full cultural flavour and significance. Again, if by cinematographic picture I could reproduce the facial expression, the bodily attitude, the significant gestures, this would add another contextual dimension.

Malinowski 1935: 26
there is the special methodological interest in the frank and full presentation of all available linguistic evidence

Malinowski 1935:4
There is no reason whatever why, in the future, an exact and physiological study of speech should not use the apparatus of sound films for reproducing fully contextualised utterances.

Malinowski 1935: 26

I'm not particularly interested in 'smoothed-over' versions of native culture. I like the stuff in the raw, as felt and dictated by the natives. [...] The genuine, difficult, confusing, primary sources. These must be presented, whatever else is done. [...] There are too many glib monographs, most of which time will show to be highly subjective performances. We need to develop in cultural anthropology that anxious respect for documentary evidence that is so familiar to the historian, the classical scholar, the Orientalist. [...] where is the raw evidence? I can't tell whether a given statement is common native knowledge or is merely your interpretation of one man's say-so.

Descriptive grammars written since Sapir and Malinowski may have been built on a corpus of recordings.

But that is not evident from the finished product as the recordings are not referenced.
Interlinear Text Archive
by J. Randolph Valentine
Department of Anthropology
University of Western Ontario
London, Ontario N6A 5C2

This stack is a very simple interlinear text archive, designed especially for use with the Rock® Grammar Authoring System.

Version Tue, Mar 25, 1997

Click mouse to continue.

First version: 1990
The traditional language documents have been the ‘Boasian’ trio of:

Grammar
Dictionary
Texts

Media - audio, fieldnotes, transcripts
- images, historical sources
- video, maps

etc ….

How to capture all of this information and keep track of it?
Fieldwork in Eratap and Erakor villages 1996–

Notes (paper)
~30 analog audio cassettes
4 hours of digital audio
3 digital video cassettes
~200 photographs
Texts based on historical sources
Maps of placenames
Transcripts (paper/computer)
Glossed texts (IGT)
Dictionary
Grammar
Language Documentation

Recent movement within linguistics, based on recognition of:

- the focus on a narrow structuralist theoretical tradition within much of linguistics that is not concerned with fieldwork

- the failure of that linguistic tradition to engage with the diversity of human languages
Language Documentation

Recent movement within linguistics, based on recognition of:

- the number of languages that may cease to be spoken in the near future
- our responsibility to create good records of the language
- developments in technology that facilitate the creation of better records
The problems that Language Documentation addresses

Language descriptions are necessarily only able to deal with some aspects of the language.

Linguistic fieldwork methods have traditionally not created records the speakers could use, but rather have been expropriative.

Language descriptions do not make explicit the links between primary records and the results of their analysis.
Problems of grammatical descriptions of languages

One volume, describing what is currently considered important information about a language
Problems of grammatical descriptions of languages

New features of language are focused on by each generation of researchers (e.g., evidentiality, referentiality, serial-verbs, switch reference ...) and older grammars do not have anything to say about them.

But grammars could be a key to analysis, assuming that there is a corpus of texts in the language.

And that the texts are accessible for others to use.
Trust me, I’m a linguist

In his 2008 memoir, “Don’t Sleep, There Are Snakes,” the linguist Dan Everett recalled the night members of the Pirahã — the isolated Amazonian hunter-gatherers he first visited as a Christian missionary in the late 1970s — tried to kill him.

Dr. Everett survived, and his life among the Pirahã, a group of several hundred living in northwest Brazil, went on mostly peacefully as he established himself as a leading scholarly authority on the group and one of a handful of outsiders to master their difficult language.

His life among his fellow linguists, however, has been far less idyllic, and Dr. Everett’s scholarship is poised to boil over anew with the ambitious new book, “Language: The Cultural Tool,” a forthcoming television documentary the scientist describes as his “most admiring view of his research among the Pirahã,” a darkly conspiratorial view of some of his colleagues.


Dr. Everett also wrote “The Cultural Tool” to argue that language should not be viewed as an innate and universal phenomenon, as is commonly held by linguists like Noam Chomsky, but as a kind of tool, built up over time by humans and useful in various ways.

In 2007 Dr. Everett published the book “Dr. Dan: The Cult of the Gift,” a collection of articles and essays that exposed the ongoing role of subversion in the politics of science.

“Don’t Sleep, There Are Snakes” was recently named one of the best books of 2008 by the National Public Radio Book Club.

Dr. Everett, who is a professor at the University of Massachusetts, is a member of the National Academy of Sciences and has been awarded the Distinguished Service Award from the American Anthropological Association.

The book is due out in May and is expected to be a major force in the ongoing debate over the nature of human language and its role in society.

Dr. Everett is a founding member of the Amazonian Language Institute, a nonprofit organization that aims to preserve endangered languages.

The book is available for purchase from Amazon.com and other online retailers.

Dr. Everett is a prolific writer and speaker and has appeared on numerous television and radio shows in the United States and abroad.

He is also the author of several other books, including “The Language Instinct,” a best-selling book that explores the origins of language and how it shapes human thought and behavior.

Dr. Everett is also the author of “The Language Myth,” a widely praised book that challenges the widely held belief that language is a natural and innate ability, and instead argues that language is a cultural artifact that is learned and developed over time.

Dr. Everett is the recipient of several awards and honors, including the prestigious Humboldt Prize, the highest academic honor given by the German government, and the prestigious MacArthur Fellowship, which is awarded to individuals who have shown outstanding creativity and potential for furthering human knowledge.

He is currently working on several new projects, including a book on the history of language, a documentary film about the language of the Pirahã, and a series of lectures on the cultural significance of language.

Dr. Everett is also the founder and president of the Amazonian Language Institute, a nonprofit organization that is dedicated to the study and preservation of endangered languages.

The institute is currently working on a number of projects, including the development of a new language documentation database, the creation of a new language laboratory, and the production of a series of language instruction videos.

Dr. Everett is also a frequent speaker and lecturer, and has appearances scheduled at several universities and conferences throughout the United States.

He is a devoted family man and is married to Elizabeth, a prominent author and scholar in her own right.

Together they have two children, a son and a daughter, who are also passionate about language and culture.

Dr. Everett is a true gentleman, a scholar of the highest caliber, and a devoted family man.

We are honored to have him on our team at Amazonian Language Institute, and we are confident that his insights and expertise will be of great benefit to our mission.

Thank you for your time, Dr. Everett.

We look forward to working with you in the future.

Sincerely,

[Signature]

Richard Henry
President, Amazonian Language Institute
Decontextualised language examples
- often the only language samples provided
- taken by others for use in new analyses
- used and reused, errors propagated
Postcolonial linguistics

Concern about

- the extraction of language records in the course of fieldwork, what useful outcomes are there for the speakers?
- responsibility to curate these records held in trust for the people recorded
Language Documentation as postmodern

Linguists creating records and situating their analysis in these records
- acknowledging the partial nature of the records
- analysis based in a corpus
- inviting others to replicate the analysis
Language Documentation

Linguists creating records and situating their analysis in these records
- emphasis on collaborations
  - with speakers
  - with other disciplines
- resulting in deeper understanding of the culture
Language Documentation

No record of a language can be complete

The earlier model in which *no* records were made is no longer tenable

Records arising from fieldwork can include any aspects of everyday life (and still be useful for structural analysis)
We want

To be able to prepare excellent data in the course of our fieldwork (without too much extra work!)

Excellent data will endure, can be accessed and can be made into various forms for delivery to various users
- Thesis / book
- DVD of examples
- Online dictionary
- Printed dictionary
- Printed collection of stories
- ITunes installation of stories
- EOPAS stories
- Archived collections of primary material
Audiamus

Created for interacting with field recordings via their transcripts

Version 1 – HyperCard – 1999
Version 2 – RunTime Revolution 2002
   LiveCode

Ipiatlik,Etasrik,Ratison. Etasrik ipiatlak mulep iseke ifcos."
1:01
There is (a natopu at) Tasiriki, the Radison. Tasiriki has a woman spirit there.
1:02
Ga me itkos. Nagien ipi Lisau.
1:02
She is there. Her name is Lisau.
1:03
Ga me itu lekor ga ega.
1:03
She looks after this place.
1:04
Rapi natopu me rutae namsor, rutae natam$ol ai naktan, lekor ptaki natam$ol.'
1:04
They are natopu (spirits) but they know people, they know the people of the village, look at them.
1:05
Tenen kin ipreg teanrun tukel, rufeikin kin tefan kin tae na ipreg namrun itakel, rufeikin kina.
1:05
Anyone who does something wrong, they show him so that he knows that he is doing something wrong, they shi.
1:06
Go natam$ol to mrotae na tenen kin ipregi, ita wi maa... [ga itu].
1:06
And the man will recognise that what he did is not good.
1:07
NT) Me ipiatlak natam$ol nen ruuter teti nanromeni?
1:07
NT) But are there people who give them some presents?
1:08
TT) Or. Or. Wel, namaroan sa nis ratam$ol, namalo$im inro saki tete natam$ol imrosaki tete aslen.
1:08
Yes. Yes. Like, bad thoughts on someone, he wishes bad things on some friend.
1:09
iset tee nanromeni ipan tua ki. Ini.
1:09
If he takes a present, he goes and gives it to her. He says, ""You will hit him for me.""
1:10
Kefo pregi taosi kin, taosi kin natam$ol nen iset nanromeni ipan tua ki.
1:10
She will do it, as that man brought a small present and gave it to her.
1:11
Malen kin natam$ol karu nen ifa tu masak, nakan kin tekan kipe pan tu natopu ki.
1:11
Then that other man might get sick, because the other one went and made an offering to the natopu a present.
1:12
Ifatu masak, panpan ale ilel na, tete murewe nen ketaa ga pam$or nlak ramsaki ga.
1:12
If he is sick, eventually he will go and see a healer (murewe) who will be able to find out for him the cause of his
1:13
Pan kefo ti na, kefo tae nrikin
1:13
Then he will tell him, ""He is the one who sooled the natopu onto you. (Lit: who gave the natopu to you)"
1:14
Go ga kefo mer tae nrikin ki na, ""P$amer sel tete nanromeni mai tao kin, me kineu kafo paa ga psi, me kafo pla.
1:14
And he would then be able to tell him, ""You now go and get a present and give it to me, and I will go and give it
1:15
Tefla. Ipiatlik tete natam$ol nen kin namaroan gar itakel, rute.
1:15
That's it. There are some people who don't think straight.
1:16
Me inalot sos$ wi man iin sos$ss.
Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC)

Collaborative digital research resource set up by the University of Sydney, University of Melbourne & Australian National University

Intermittent funding from the Australian Research Council
PARADISEC

The Pacific and Regional Archive for Digital Sources in Endangered Cultures offers a facility for digital conservation and access to endangered materials from all over the world. Our research group has developed models to ensure that the archive can provide access to interested communities and conforms to emerging international standards for digital archiving. A primary motivation for this project is making field recordings available to those recorded and their descendants. While initially conceived of as an archiving project focusing on the Asia-Pacific region, PARADISEC has developed into a centre for activities including:

- **Training** in data management (lexicographic software and transcription and interlinearising of texts), recording techniques and data linkage
- **Provision of a catalogue** in which users can create descriptions of their collections
- **Standardized Metadata** - creating the descriptions in forms that conform to relevant standards and are harvested by the Open Archives Initiative search engines, enabling access for a wider community
- **Building models** that show how to make reusable data (e.g. ExSite9) and using current tools such as Elan and Toolbox
- **Building models** that show how to reuse data (e.g. EOPAS, online dictionaries, ITunes repositories)
- **Cultural preservation** – backup and data provision for cultural agencies in the region (e.g. Institute for PNG Studies, Vanuatu Cultural Centre)
- **Global Focus** – meeting the demand to house files with a research focus from all over the world
PARADISEC goals

Digital archiving of endangered recorded field material from the region around Australia and more broadly

Exploit potential of digital systems to build a collaborative cross-institutional resource

Develop and implement electronic management of a digital research archive

Create linkages between Australian research institutions, national archival institutions, stakeholders in the Asia-Pacific region, and international bodies to ensure future viability of the resource
Records not being looked after
Typical legacy data – the original motivation

Stephen Wurm’s several hundred tapes, including 120 1970s Solomon Islands tapes and transcripts/fieldnotes

Arthur Capell’s 114 tapes, Pacific and PNG 1950s (and 30 archive boxes of fieldnotes)

Bert Voorhoeve’s 180 tapes - West Papua

Tom Dutton’s 295 PNG tapes
Anthropological recordings

Geoff White, Cheke Holo, Solomon Islands
Don Gardner, Mian, PNG
Lamont Lindstrom, Tanna, Vanuatu
Ted Schwartz, Manus Island, PNG
Andrew McIntyre, Vanuatu
Ted Schwartz, PNG

- See the catalog
Legacy material

“an estimated 80% of the world-wide holdings representing the cultural and linguistic diversity of mankind are not held by audiovisual archives proper.”

Dietrich Schüller 2004: 9
2003 – Australian Research Council Linkage Infrastructure and Equipment Fund grant to establish a digitisation suite in Sydney

- Located, described and digitised ~500 hours of analog audio tapes

Established digitisation units at ANU and Melbourne

Established mass data storage system

paradisec.org.au
Storage

2004 - Australian Partnership for Advanced Computing (APAC)

2009 - National Computational Infrastructure (NCI)

2012 - Research Data Storage Initiative (RDSI)
Preservation

EBU Broadcast Wave Format (BWF), Quadriga

International standard (developed by European Broadcast Union) adopted by Australian national institutions (National Film and Sound Archive, National Library of Australia, AIATSIS, etc)

24-bit, 96khz uncompressed PCM audio

Encapsulated metadata includes unique permanent identifiers, coding history, and content descriptors

Digital ‘sealing’ for data authentication provided by Quadriga Audio-Cube system
Collection metrics – September 2013

757 languages
249 collections
8,896 items
56,114 files
7,644 hours (of both wav and derived mp3)
6.73 TB of files

903 GB of video files in archive
5.38 TB of audio files in archive
117 GB of image files in archive
130 MB of text files in archive
Funding – flood and drought

![Graph showing funding over years]
Catalog

Conforms to relevant standards

  Dublin Core, Open Archives Initiative, Open Language Archives Community

Initial Filemaker Pro version

- controlled vocabularies via popup menus
- export to XML for backup and online access

Second version in SQL/PHP
Catalog has language identifiers (ISO 639-3)

Allows users to enter the local name even if there is no ISO-639-3 code

Geographic information (DC bounding box)

All maximise locatability of the item

Often the use of the catalog is the first time the user has organised their
PARADISEC database version 1 (FMPro)
PARADISEC database version 3 (Ruby)

PARADISEC Catalog

Item details
- Item ID: NTB-TEXT
  - Title: Dictionary and texts in South Efate
  - Description: A dictionary and texts in South Efate. Created in Toolbox, this is a work-in-progress that will be updated periodically. Toolbox requires settings files (TYP) that accompany the main data. Thus the file NTB-TEXT-DICTYP_[DATE].bt has the accompanying file NTB-TEXT-DICTYP_[DATE].bt. The collection of texts in NTB-TEXT-TXT_[DATE].bt has the TYP file in NTB-TEXT-TXTYP_[DATE].bt.
  - Origin date: 2006-11-05
  - Archive link: http://catalog.paradisec.org.au/repository/NTB/TEXT
  - Collector: Nick Thieberger
  - Countries: Vanuatu - VU
  - Language as given: Efate, South - erk
  - Subject language(s): Efate, South - erk

Content Files (13)

<table>
<thead>
<tr>
<th>Filename</th>
<th>Type</th>
<th>File size</th>
<th>Duration</th>
<th>Metadata</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTB-TEXT-DICTYP_20060904.bt</td>
<td>text/plain</td>
<td>13.6 KB</td>
<td>View</td>
<td></td>
</tr>
<tr>
<td>NTB-TEXT-DICTYP_20070827.bt</td>
<td>text/plain</td>
<td>14.5 KB</td>
<td>View</td>
<td></td>
</tr>
<tr>
<td>NTB-TEXT-DICTYP_20110501.bt</td>
<td>text/plain</td>
<td>14 KB</td>
<td>View</td>
<td></td>
</tr>
<tr>
<td>NTB-TEXT-DICTYP_20061004.bt</td>
<td>text/plain</td>
<td>57 KB</td>
<td>View</td>
<td></td>
</tr>
<tr>
<td>NTB-TEXT-DICTYP_20070827.bt</td>
<td>text/plain</td>
<td>595 KB</td>
<td>View</td>
<td></td>
</tr>
<tr>
<td>NTB-TEXT-DICTYP_20091130.bt</td>
<td>text/plain</td>
<td>602 KB</td>
<td>View</td>
<td></td>
</tr>
<tr>
<td>NTB-TEXT-DICTYP_20110501.bt</td>
<td>text/plain</td>
<td>619 KB</td>
<td>View</td>
<td></td>
</tr>
<tr>
<td>NTB-TEXT-TXTY_20060904.bt</td>
<td>text/plain</td>
<td>5.71 KB</td>
<td>View</td>
<td></td>
</tr>
<tr>
<td>NTB-TEXT-TXTY_20110501.bt</td>
<td>text/plain</td>
<td>6.27 KB</td>
<td>View</td>
<td></td>
</tr>
<tr>
<td>NTB-TEXT-TXT_20061004.bt</td>
<td>text/plain</td>
<td>919 KB</td>
<td>View</td>
<td></td>
</tr>
<tr>
<td>NTB-TEXT-TXT_20070827.bt</td>
<td>text/plain</td>
<td>999 KB</td>
<td>View</td>
<td></td>
</tr>
<tr>
<td>NTB-TEXT-TXT_20091130.bt</td>
<td>text/plain</td>
<td>1.05 MB</td>
<td>View</td>
<td></td>
</tr>
<tr>
<td>NTB-TEXT-TXT_20110501.bt</td>
<td>text/plain</td>
<td>1.13 MB</td>
<td>View</td>
<td></td>
</tr>
</tbody>
</table>

13 files total, 6.44 MB

Collection Information
- Collection ID: NTB
- Collection title: Dictionary of South Efate
Viewing our catalog

Welcome to the catalog of the PARADISEC collection

http://catalog.paradisec.org.au
Viewing our catalog
- via the Australian National Data Service

http://researchdata.ands.org.au
Viewing our catalog
- via Virtual Language Observatory

http://catalog.clarin.eu/vlo
Viewing our catalog
- via WorldCat

https://www.worldcat.org/
Viewing our catalog
– via the Open Language Archives Community

OLAC resources in and about the Mae language

ISO 639-3: mme

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

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

Other known names and dialect names: Dirak, North Small Nambas

Use faceted search to explore resources for Mae language.

Primary texts

1. Online: Tirax Field Tape 001. Amanda Brotchie (compiler); Frank Davey (depositor); Amanda Brotchie (recorder). 2004. Pacific And Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). oai.paradiseq.org.au:AB1-001
2. Online: Tirax Field Tape 002. Amanda Brotchie (compiler); Frank Davey (depositor); Amanda Brotchie (recorder). 2004. Pacific And Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). oai.paradiseq.org.au:AB1-002
3. Online: Tirax Field Tape 005. Amanda Brotchie (compiler); Frank Davey (depositor); Amanda Brotchie (recorder). 2004. Pacific And Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). oai.paradiseq.org.au:AB1-005
4. Online: Tirax Field Tape 006. Amanda Brotchie (compiler); Frank Davey (depositor); Amanda Brotchie (recorder). 2004. Pacific And Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). oai.paradiseq.org.au:AB1-006
5. Online: Tirax Field Tape 009. Amanda Brotchie (compiler); Frank Davey (depositor); Amanda Brotchie (recorder). 2004. Pacific And Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). oai.paradiseq.org.au:AB1-009
6. Online: Tirax Field Tape 010. Amanda Brotchie (compiler); Frank Davey (depositor); Amanda Brotchie (recorder). 2004. Pacific And Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). oai.paradiseq.org.au:AB1-010
7. Online: Tirax Field Tape 013. Amanda Brotchie (compiler); Frank Davey (depositor); Amanda Brotchie (recorder). 2004. Pacific And Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). oai.paradiseq.org.au:AB1-013
8. Online: Tirax Field Tape 018. Amanda Brotchie (compiler); Frank Davey (depositor); Amanda Brotchie (recorder). 2004. Pacific And Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). oai.paradiseq.org.au:AB1-018
9. Online: Tirax Field Tape 019. Amanda Brotchie (compiler); Frank Davey (depositor); Amanda Brotchie (recorder). 2004. Pacific And Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). oai.paradiseq.org.au:AB1-019
10. Online: Tirax Field Tape 020. Amanda Brotchie (compiler); Frank Davey (depositor); Amanda Brotchie (recorder). 2004. Pacific And Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). oai.paradiseq.org.au:AB1-020
11. Online: Tirax Field Tape 021. Amanda Brotchie (compiler); Frank Davey (depositor); Amanda Brotchie (recorder). 2004. Pacific And Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). oai.paradiseq.org.au:AB1-021
Documentation index, Vanuatu languages

123 languages: Grammar 1-5, Lexicon 1-5, Texts 1-5, Media corpus 1-5
Arthur Capell (1902-1986), Guide to Records

Listed by Peter Newton, Austhec and PARADISEC

Arthur Capell was an Australian linguist and ethnographer who spent much time recording and documenting both Australian Aboriginal languages and endangered languages in the Asia - Pacific region. The Arthur Capell textual collection consists of some thirty boxes of notes, transcripts, and other supporting materials. What is documented in this guide are some of the records relating to his non-Australian linguistic work that have been digitally imaged.

Please do not copy material from this site for further distribution but rather link to this site. PARADISEC has raised funds to digitise this collection and would like to be recognised for the work that we have put into developing the online presentation of fieldnotes. If you copy and distribute this data and do not acknowledge PARADISEC's work then we will have to put password protection on the data.

Copyright: Paradisec believes that many of the items provided through this guide are no longer the subject of copyright restrictions, or have been cleared for display in this service by the Copyright owners. However, Paradisec invites any individuals who believe they hold current rights over items provided through this service to make contact.

[Details]

- About the records
  - Scope and content
  - How to use this finding aid
  - Archival terms

- Provenance - creators and former custodians

- Series list and summary descriptions

- Inventory listing by series
  - Series 01 - Personal and Biographical Material
  - Series 02 - General Linguistic and Ethnological Materials
  - Series 03 - Indonesia - Regions and Languages
  - Series 04 - Bougainville, Buka and Offshore Islands - Regions and Languages
  - Series 05 - Melanesia and Oceania - Regions and Languages
  - Series 06 - Solomon Islands - Regions and Languages
  - Series 07 - Timor - Regions and Languages
  - Series 08 - Vanuatu - Regions and Languages

Arthur Capell (1902-1986), Guide to Records
Item: VEFAT25
Contact PARADISEC
To print this page you may need to select landscape paper orientation for better results.

Natus ni Natuanausen

From Pastor Sope (Aged 82 years) Written in Earakor Language

Natuanausen nig Mantiki.

Naur lab nig Melanesian rumani til natuanausen nig Mantiki, tete ru til Manu ike, go tili ru til Manu irru, go tete ru til i til Manu i tot. Manu on siin nig Man i tot i titimori i bi Man i tot, nalaken Manun ru bi natamol nag ru tok umromen, seluase nobu ki teps umromen rarui nig. Nua ki. 
DVDs of field recordings archived directly from the northern Philippines

Archiving before analysis

Citation of primary data in the analysis
Linkages

Importance of relationships with regional cultural organisations, including repatriation of CD copies of tapes, and digitisation of reel-to-reel tapes

- Vanuatu Kaljoral Senta
  - provision of safe ‘blind’ backup of their digitised sound collection
- Institute of PNG studies
- Rapanui community

News/Comment
From p1... Culture

The collaborative efforts have already paid off with the presentation to the Cultural Centre’s Ralph Regenvanu and Ambong Thompson of valuable sound recordings of Nagriamel by Nick Thieberger of the Australia based Pacific & Regional Archive. On page 1 the women field workers workshop participants are at work.
Digitisation of mouldy tapes in Drehu for Leonard Drilë Sam at the University of New Caledonia
Repatriation

Erakor school, Efate, Vanuatu
**Making thatch ; Efate, South (erk) ; Vanuatu**

### Set Displayed Tracks
- **Malen amurin na katur rowat,**
  - *Malen amurin* na katur rowat,
    - malen *a=*  
      - *mur-*i  
      - *na ka=*  
      - *tur rowat*
    - as 1S.RS=want  
      - TS 3S.O COMP 1S.IRS=sew  
      - sago_palm
  - *When I want to sew thatch,*
  - go apo pan slat rowat, kafan slat rowat.
  - go  
    - apo  
    - pan slat rowat  
    - kafan slat rowat.
  - and 1S.RS=PSP go  
    - take sago_palm 1S.IRS=go:IR take sago_palm
  - *and I will get sago palm leaves (rowat), I'll get rowat.*
  - Aler mai, kafo pei ṛhasil rowat, aṕhasil rowat inom.
  - Aler  
    - mai  
    - kafo  
    - pei ṛhasil rowat  
    - aṕhasil rowat  
    - inom.
  - 1S.RS=return come 1S.IRS=PSP:IR first derib  
    - sago_palm 1S.RS=derib sago_palm 3S.RS=finish
  - *I come back, I'll prepare the rowat, prepare it until it's done.*
  - Amer ṕelkia, aṕelkin itu.
  - Amer  
    - ṕelkia  
    - aṕelkin  
    - itu.
  - 1S.RS=again bend  
    - TR 3S.O 1S.RS=bend  
    - TR 3S.O 3S.RS=stand
  - *Then I'll fold it, it is folded.*
  - Me apo mer pan tai lop. Apan tai lop, mai. Apregtaki, taipetakio
  - Me apo  
    - mer pan tai lop  
    - Apan tai lop  
    - mai.
  - me  
    - apo  
    - mer pan tai lop  
    - a=pan tai lop mai
  - but 1S.RS=PSP again go  
    - cut bamboo 1S.RS=go cut bamboo come
  - Apregtaki,  
    - taipetakio
  - a= 
    - preg  
    - pta  
    - tai pta  
    - ki -ø
  - 1S.RS=make make  
    - good  
    - TR cut make  
    - good  
    - TR 3S.O
  - *I get it ready, cut it to make it ready. Then I'll cut bamboo. I cut bamboo and come.*
  - Ileg nen kin kafo tur rowat, apreg, atonkin rupitkaskei. Lop rupitkaskei.
  - Ileg  
    - nen kin  
    - kafo tur rowat  
    - apreg,
  - =  
    - leg  
    - nen kin  
    - ka=fo tur rowat
  - a=preg
  - 3S.RS=straight that REL 1S.IRS=PSP:IR sew sago_palm 1S.RS=make
  - atonkin  
    - rupitkaskei. Lop  
    - rupitkaskei.
  - a=  
    - ton  
    - ki -n  
    - ru=  
    - pitkaskei lop  
    - ru=  
    - pitkaskei
Selection of XML formats to upload
Validation of XML on upload
Ruto los ena, ruto pan los Ermag me gar,
3P.RS= PROG wash LOC- hesit 3P.RS= PROG
The washed there, they would go and wash at Erromango, but they,
Import Transcripts and media
Media upload and transcode
<table>
<thead>
<tr>
<th>Date</th>
<th>Depositor</th>
<th>Title</th>
<th>Description</th>
<th>Language</th>
<th>Country</th>
<th>Media</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>1966-01-31</td>
<td>Nick Thieberger</td>
<td>Abiamio Nakikai 57 phrases</td>
<td>A story recorded by John Harris in Kiwai, from PNG in the mid 1960s.</td>
<td>kiw</td>
<td>PG</td>
<td>Kiwai recording</td>
<td>Show/Editor/ Delete</td>
</tr>
<tr>
<td>2000-04-07</td>
<td>Nick Thieberger</td>
<td>Angels and Erromango 90 phrases</td>
<td>A story about a woman who flies to Efate and whose wings are stolen by a man who wants her for his wife. Her children eventually find her wings and return them to her and she flies home.</td>
<td>erk</td>
<td>VU</td>
<td>20003b</td>
<td>Show/Editor/ Delete</td>
</tr>
<tr>
<td>1998-10-03</td>
<td>Nick Thieberger</td>
<td>Asaraf 14 phrases</td>
<td>Asaraf was a giant who could walk from Efate to Erromango and who pushed Erromango away from Efate.</td>
<td>erk</td>
<td>VU</td>
<td>98007A</td>
<td>Show/Editor/ Delete</td>
</tr>
<tr>
<td>2011-04-11</td>
<td>Nick Thieberger</td>
<td>CP Drowning 58 phrases</td>
<td></td>
<td>mwf</td>
<td>AU</td>
<td>9-08-CP-drowningMurrinhPatha</td>
<td>Show/Editor/ Delete</td>
</tr>
<tr>
<td>1998-10-20</td>
<td>Nick Thieberger</td>
<td>Daisy Wayane, A life story 54 phrases</td>
<td>Daisy Wayane talking about the various jobs she has had in her life.</td>
<td>erk</td>
<td>VU</td>
<td>98010-B</td>
<td>Show/Editor/ Delete</td>
</tr>
<tr>
<td>1998-09-29</td>
<td>Nick Thieberger</td>
<td>Darkness, light and Christianity 4 phrases</td>
<td>A brief description of the coming of Christianity</td>
<td>erk</td>
<td>VU</td>
<td>98003A</td>
<td>Show/Editor/ Delete</td>
</tr>
<tr>
<td>1996-03-07</td>
<td>Nick Thieberger</td>
<td>English Police 30 phrases</td>
<td>Kalsarap talking about his time with the English Police in the 1930s. He is reading from his notes, and made some slight edits to the transcript.</td>
<td>erk</td>
<td>VU</td>
<td>NT1-004-B</td>
<td>Show/Editor/ Delete</td>
</tr>
<tr>
<td>2011-04-12</td>
<td>Nick Thieberger</td>
<td>Harry Laoto’s story 21 phrases</td>
<td></td>
<td>erk</td>
<td>VU</td>
<td>98010-B</td>
<td>Show/Editor/ Delete</td>
</tr>
<tr>
<td>1998-09-29</td>
<td>Nick Thieberger</td>
<td>How girls can behave 22 phrases</td>
<td>A description of how girls used to be expected to behave.</td>
<td>erk</td>
<td>VU</td>
<td>98003b</td>
<td>Show/Editor/ Delete</td>
</tr>
<tr>
<td>Date</td>
<td>Depositor</td>
<td>Title</td>
<td>Description</td>
<td>Language</td>
<td>Country</td>
<td>Media</td>
<td>Action</td>
</tr>
<tr>
<td>------------</td>
<td>------------</td>
<td>----------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------</td>
<td>---------</td>
<td>-----------</td>
<td>------------</td>
</tr>
<tr>
<td>2000-04-07</td>
<td>Nick Thieberger</td>
<td>Angels and Erromango 90 phrases</td>
<td>A story about a woman who flies to Efate and whose wings are stolen by a man who wants her for his wife. Her children eventually find her wings and return them to her and she flies home.</td>
<td>erk</td>
<td>VU</td>
<td>20003b</td>
<td>Show</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Asaraf 14 phrases</td>
<td>Asaraf was a giant who could walk from Efate to Erromango and who pushed Erromango away from Efate.</td>
<td>erk</td>
<td>VU</td>
<td>98007A</td>
<td>Edit</td>
</tr>
<tr>
<td>1998-10-20</td>
<td>Nick Thieberger</td>
<td>Daisy Wayane, A life story 54 phrases</td>
<td>Daisy Wayane talking about the various jobs she has had in her life.</td>
<td>erk</td>
<td>VU</td>
<td>98010-B</td>
<td>Show</td>
</tr>
<tr>
<td>1998-09-29</td>
<td>Nick Thieberger</td>
<td>Darkness, light and Christianity 4 phrases</td>
<td>A brief description of the coming of Christianity</td>
<td>erk</td>
<td>VU</td>
<td>98003A</td>
<td>Edit</td>
</tr>
<tr>
<td>1996-03-07</td>
<td>Nick Thieberger</td>
<td>English Police 30 phrases</td>
<td>Kalsarap talking about his time with the English Police in the 1930s. He is reading from his notes, and made some slight edits to the transcript.</td>
<td>erk</td>
<td>VU</td>
<td>NT1-004-B</td>
<td>Show</td>
</tr>
<tr>
<td>2011-04-12</td>
<td>Nick Thieberger</td>
<td>Harry Laoto’s story 21 phrases</td>
<td></td>
<td>erk</td>
<td>VU</td>
<td>98010-B</td>
<td>Delete</td>
</tr>
<tr>
<td>1998-09-29</td>
<td>Nick Thieberger</td>
<td>How girls can behave 22 phrases</td>
<td>A description of how girls used to be expected to behave.</td>
<td>erk</td>
<td>VU</td>
<td>98003b</td>
<td>Show</td>
</tr>
<tr>
<td>1998-09-30</td>
<td>Nick Thieberger</td>
<td>Independence 26 phrases</td>
<td>Discussion of the time of Independence in Vanuatu.</td>
<td>erk</td>
<td>VU</td>
<td>98004a</td>
<td>Edit</td>
</tr>
<tr>
<td>1998-10-23</td>
<td>Nick Thieberger</td>
<td>Katapel and Liportani 25 phrases</td>
<td></td>
<td>erk</td>
<td>VU</td>
<td>98009A</td>
<td>Delete</td>
</tr>
</tbody>
</table>
Making thatch; Efate, South (erk); Vanuatu

Playable media

Metadata

Unlink Media Item
Delete transcript
Reference to morpheme-level

Keyword in Context / Concordance in all texts of that language

Selected text
Ability to turn off morphemic view
Reference to timed chunk
EOPAS – text and media corpus presentation

Upload text in interlinear format

Upload media

Present text and media online

Citable text and media

www.eopas.org

Planned

Allow media in PARADISEC to be called directly
Problem of data collected without sufficient metadata

Reluctance of researchers to create metadata

Make it easier by providing tools
Laptop organisation of collections – ExSite9

http://youtu.be/4_NmUXrWZIE

Digital Endangered Languages and Musics
Archives Network - delaman.org

DoBeS LAT Archive
dobes.mpi.nl

Endangered Languages Archive
www.hrelp.org/archive

Projet Archivage
lacito.vjf.cnrs.fr/archivage

The LACITO Archive
An archive of natural speech in "rare" languages
The LACITO Archive provides free access to documents of indigenous, semi-official, regional, dialectal, and local languages spoken in France and its overseas territories. The archive is a resource for researchers, educators, and anyone interested in language diversity.

Copyright
The archive is the result of a project supported by the French National Center for Scientific Research. Reproduction of all or part of the archive is prohibited without the express consent of the author.

Archives Network - delaman.org

Alaska Native Language Center
www.uaf.edu/anlc

Sam Noble Museum, University of Oklahoma
www.snomnh.ou.edu/collections-research/nal.htm

The Rosetta Project
rosettaproject.org

California Language Archive
cla.berkeley.edu

Archive of the Indigenous Languages of Latin America
www.ailla.utexas.org

Kaipuleohone, U.Hawaii
www.ling.hawaii.edu/langdoc/archive.html

PARADISEC
paradise.org.au

The Rosetta Project
The Rosetta Project is a global collaboration of language specialists and native speaking communities working to build a publicly accessible digital library of human languages.
Archiving will become normal practice for linguists

LSA resolution 2010

‘Linguistic Society of America supports the recognition of these materials as scholarly contributions to be given weight in the awarding of advanced degrees and in decisions on hiring, tenure, and promotion of faculty. It supports the development of appropriate means of review of such works so that their functionality, import, and scope can be assessed relative to other language resources and to more traditional publications.’
RNLD aims to advance the sustainability of indigenous languages and to increase the participation of Indigenous peoples in all aspects of language documentation and revitalisation through training, resource sharing, networking, and advocacy.

What we provide

- Networking — we facilitate international networking through our active email discussion list and Facebook group, support for the Transient Languages blog, and links to other blogs and networks. Archived RNLD discussions are publicly accessible and searchable through the Linguist List.

- Resources — we disseminate information about conferences, funding opportunities, education and training programs, policies, publications, film, radio and television, news, and the methods and technologies needed to document, archive, revitalise and maintain indigenous languages.

- Training — we partner with Indigenous communities and organisations to develop capacity in language documentation and revitalisation. Our flexible training materials and practices ensure that the skills which people gain are used, retained and shared more widely within a community.

- Train-the-trainer — we promote autonomy and sustainability through the Documenting and Revitalising Indigenous Languages program, which mentors skilled language workers in becoming trainers in linguistic documentation and revitalisation methods within their own communities and organisations.

- Advocacy — we advocate on issues such as bilingual education, indigenous literacy, language policies and recognition for Indigenous peoples, raise awareness through public events on occasions such as International Mother Language Day, and organize informal gatherings for linguists and community language activists.

Need for training and advice
Methods for conducting research
Data management
Appropriate tools for analysis and exploration of data

http://rnld.org
Conclusions

Language Documentation acknowledges the partiality of the data on which an analysis is based.

It explicitly inserts the linguist into the process of recording, annotating and preparing the corpus for scrutiny by others.

It emphasises the presentation of the context of an utterance.
Conclusions

Language Documentation offers a scientific method for replication of analyses, external review of collections and processes, and data recording and sharing.

PARADISEC as an example of humanities scholars taking responsibility for building the necessary infrastructure to curate research outputs and primary materials.
The speaker’s attendance at this conference was sponsored by the Alexander von Humboldt Foundation.

http://www.humboldt-foundation.de
Structuring legacy material

Identifying legacy material

Manuscripts as major source of historical information
Daisy Bates vocabularies

24,700 page images, unknown number of languages

Mixed types of material

- Handwritten manuscripts
  - Questionnaires (based on a template) (13,487 pp)
  - Freeform lists (1,251 manuscript pages)

- Typescripts
  - Of questionnaires (4,933 pp)
  - Of freeform lists (some related to handwritten forms, but some not)
500 copies of a blank wordlist questionnaire produced by the Government Printer in Perth in 1904

Around 220 filled out
General locations of vocabularies (where known). Note that both manuscript and typescript versions of the same vocabulary are referenced here.

Kulaji, Ijala, Warduna, Eyre, MSS.
W. Williams, Eucia and Eyres Sandpatch, MSS.
Dr Wilson–K.G. Sound, MSS.
Captain King–K.G. Sound, MSS.
Dr Scott Nind–K.G. Sound, MSS.
Compiled by various contributors King George's Sound, Various contributors–King Georges Sound, MSS.
Jakumb and Wabbinyet of Albany.
Boordenam, Beereenan, Wardabirt and Boongong of Albany.
Woorgomirt of Albany.
Boordenam, Beereenan, Wardabirt and Boongong of Albany. MSS.
Woorgomirt of Albany. MSS.
Leaflet no. 4. Albany, K.G.S.
Leaflet no. 4. Albany, K.G.S. MSS.
Vocabulary from Albany and Denmark.
Vocabulary from Albany and Denmark. MSS.
Watenbury and Banyaich of the Plantageret district.
Watenbury and Banyaich of the Plantageret district. MSS.
Michael Quinn–Katanning, John Moir–Plantagenet L.P. and G. Hall–Hay and Murray districts Leaflet no. 4–Albany
Leaflet no. 1–King Georges Sound, Dr Wilson, Captain King, Dr Scott Nind, Captain Flinders, – K. G. Sound, Geo Chester, R. T. Goldworthy
As above, MSS.
John Moir–Plantagenet dist. MSS.
Bardeet of Bridgetown.
Ngalbaich of Korrup, Bardeet of Bridgetown. MSS.
Ngalbaich of Jerramungup, Ngalbaich of Jerramungup. MSS.
John L. Wall, Jerramungup natives.

Man, his relationships, etc.

Aunt

Baby

Blackfellow

Unique identifier assigned to each item in the questionnaire

Language identifier (ISO-639 where possible) assigned to each item in the questionnaire

Personal name is marked to allow it to be listed as a name

Geographic location marked to allow geo-coding

Page number of questionnaire allows links from typescript
Fish (generic)         n-ar-da
Oyster (large Pearl)  <add> Oy-
frog (generic)         ma-lee
Iguana, long-tailed   wongalee
Lizard (generic)      boon-jar-da
Snake (generic)       warl-la
Ant (common small)    bing-a
Ant (generic)         biyu
Ant (white)           muyo
Afternoon             beer-inga
Cold                  mool-tha
Country (open)        bulgarra
Country (stony)       munda-mi
Creek                 moorie
Dark                  <add> Dark
Floods                cu-r-ry-ac-ca
Grass                 min-na
Hole                  mar-do
Moon                  weel-arra
Morning               wunmunda
Pool in river         babbaye
Rain                  yoong-o
River                 wandi

Fish (generic)         n-ar-da
Oyster                pirn pwer
Frog (generic)         ma-lee
Iguana, long tailed   wongalee
Lizard (generic)      boon-jar-da
Snake (generic)       warl-la
Ant, common, small    bing-a
Ant (generic)         biyu
Ant (white)           muyo
Afternoon             beer-inga
Cold                  mool-tha
Country (open)        bulgarra
Country (stony)       munda-mi
Creek                 moorie
Dark                  <add> Dark
Floods                cu-r-ry-ac-ca
Grass                 min-na
Hole                  mar-do
Moon                  weel-arra
Morning               wunmunda
Pool in river         babbaye
Rain                  yoong-o
River                 wandi

Water                   baba
Water (deep)             cannanse
Water (running)          carraka

Reptiles.

Frog, edible          ...
Frog, large green     ...
Frog, gray green...
Frog, long-tailed...
Iguana, green...
Iguana, long tailed...
Iguana, stumpy...
Lizard (generic)      boon-jar-da
Sea-soup               warl-la
Snake (generic)       wongalee
Turtle (long-necked, fresh water)