



the 'reach' of digital language archives: towards criteria for evaluation

David Nathan



Endangered Languages Archive
SOAS, University of London
www.elar-archive.org

Paradisec RRR 2013
University of Melbourne
2 December 2013



Aims



- initial suggestions towards a set of criteria for considering and evaluating the “reach” (access, accessibility, awareness) of our archives
- complementary to other crucial areas such as preservation etc
- approaches to “reach” could be much more scientific than they are



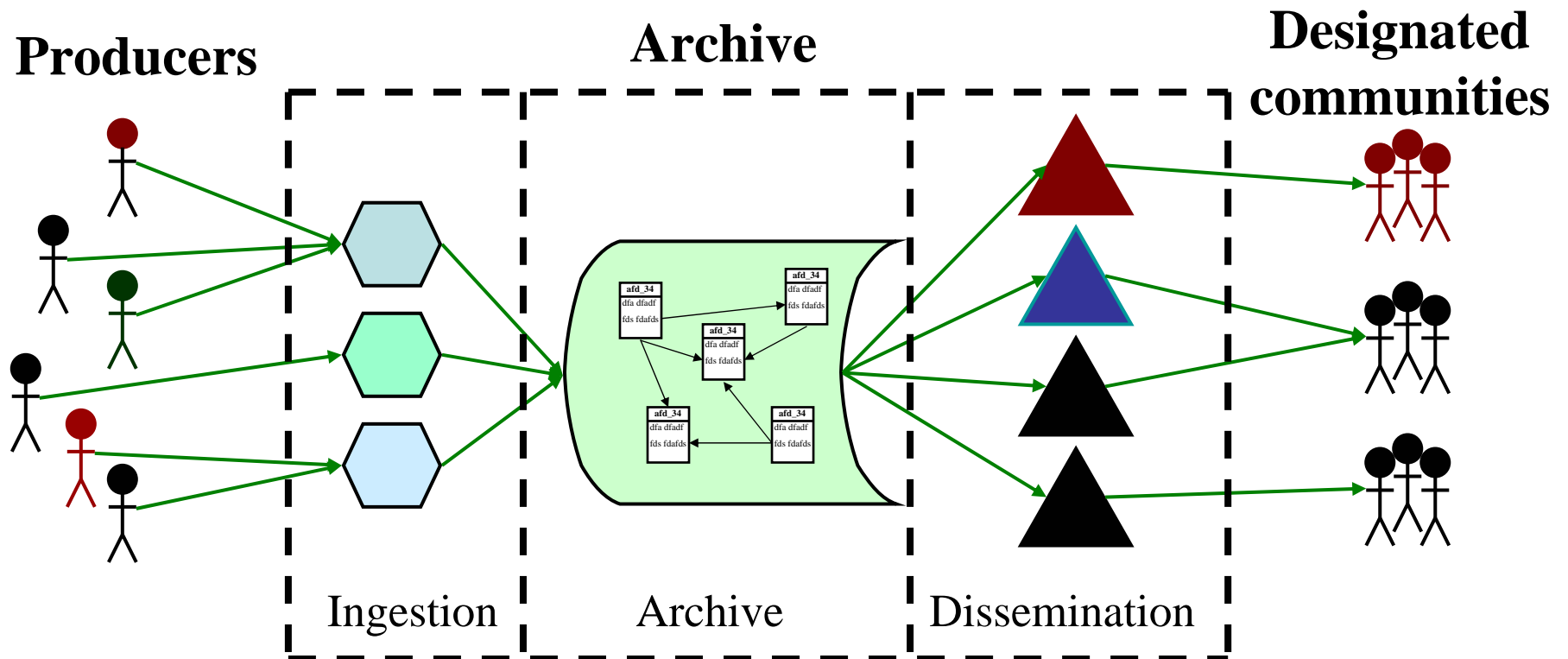
Archive evaluation schemes



- Open Archival Information Systems OAIS (2002)
- NINCH Guide (2002)
National Initiative for a Networked Cultural Heritage
- Data Seal of Approval (2009)
- TAPS (2010)

OAIS reference model

- 3 'packages' *ingestion, archive, dissemination*
- fairly one way flow
- recognise communities of producers and users





Archive evaluation schemes



- Open Archival Information Systems OAIS (2002)
- NINCH Guide (2002)
National Initiative for a Networked Cultural Heritage
- Data Seal of Approval (2009)
- TAPS (2010)



TAPS



- TAPS (Target, Access, Preservation, and Sustainability, Chang 2010) addresses:
 - recommended best practices for the long-term preservation of digital information (preservation)
 - areas of special concern to linguists and language communities (access)

TAPS on “reach”

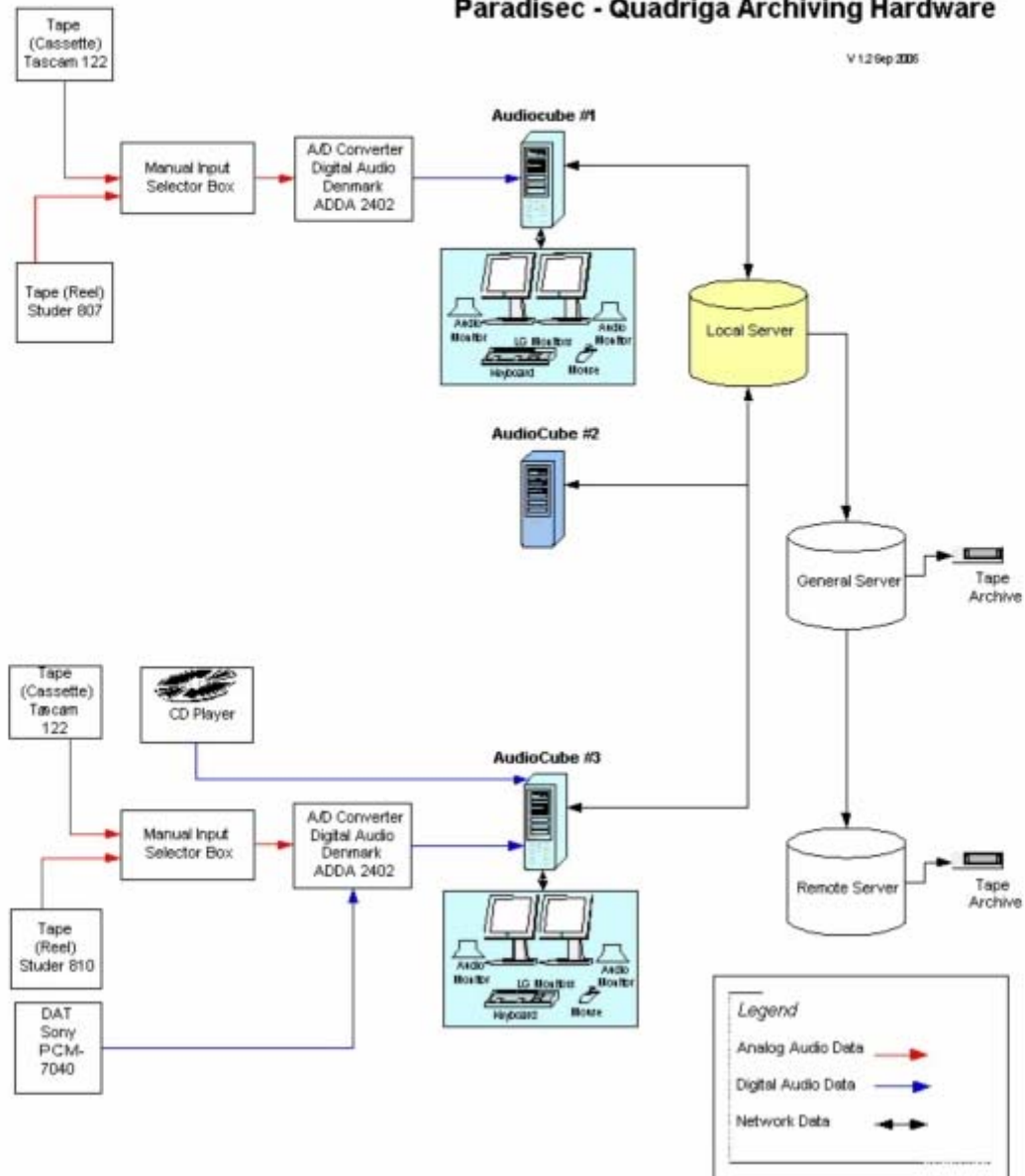
- Will the audience (*designated communities*) that I wish to reach be able to access archive materials?
- Will members of the be expected to have access to the Internet?
- Will they need to maintain an email address?
- Will the metadata be available in English only, or will it be available in another language that is more accessible to them?
- Will the archive charge fees for copies of data on media that are usable by them?

“Reach” 1

- *acquisition*: policy and needs-driven acquisition, ingest and curation, increases coverage

Paradisec - Quadriga Archiving Hardware

V 1.2 Sep 2005



“Reach” 2

- *audiences*: understanding of audiences to provide appropriate services for them, e.g. their languages of access, their varied technological and information literacies, interface design and usability



El Archivo de los Idiomas Indígenas de Latinoamérica

Idiomas

English ▾ Vaya

Registro/Ingreso

Bienvenido

Cómo usar

Navegar

Depositantes

Derechos

Idiomas

Eventos

Presentaciones de
AILLA

Contáctenos

Ayuda

Enlaces

Colección: Southern Zapotec Languages Collection

[Visualizar los recursos en esta colección](#)

Título	
Título español	Colección de Lenguas Zapotecas de la Sierra Sur
Título inglés	Southern Zapotec Languages Collection
Lenguas en esta colección:	Zapotec, Coatec, Zapotec, Coatec, Zapotec, Isthmus, Zapotec, Miahuatec, Zapotec, Sierra Juárez
Coleccionista(s)	Rosemary Beam de Azcona
Depositante(s)	Rosemary Beam de Azcona
Sitio del Proyecto/Coleccionista	
Descripción	Esta colección incluye grabaciones de audio en varias variedades de zapoteco habladas en Oaxaca, México y textos manuscritos. Las grabaciones representan una gran variedad de géneros de discurso. Los textos incluyen notas del campo, juegos de datos, y transcripciones y traducciones de textos grabados.
Referencias	

Resumen de los contenidos de la colección

Géneros	Lista de palabras; Encuentro; Instrucciones; Historia; Juego de datos; Entrevista; Gramática; Léxico; Narrativa; Descripción; Comentario; Esbozo; Procedimiento; Conversación; Juego de habla		
Número de archivos archivales	1046	Porcentaje de archivos restringidos	14
Número de grabaciones	46	Duración total de audio	15:37:57

Pingjiang traditional love songs

[Home](#) [Resources](#) [Comments](#)

This deposit has been accessioned.

Search this deposit

[Reset keywords](#)

Access protocol

 (77)

Language

 Chinese (69)
Pingjiang (70)

Type

 Audio (69)
Image (6)
Spreadsheet (1)
Text (1)

Tags

Metadata (1)

Genre

 Love song (69)
Song lyrics (1)
Song (69)

 Participants [more](#)

 Chen Cuihua (54)
Li Shuguang (1)
Mao Chuan (8)
Wang Xunfu (1)
Zhan Miaochun (1)

Pingjiang traditional love songs

Language: Pingjiang

Depositor: Shenkai Zhang

Location: China



Group represented

Pingjiang (Han Chinese)

Language information

Pingjiang dialect is spoken in Pingjiang County, in the northeast of Hunan province and borders with Hubei and Jiangxi provinces. See also "Other information". 平江方言是湖南省平江县内的方言。平江位于湖南东北部、是湖南、湖北、江西的三省交界之处。

Deposit contents

The Pingjiang Traditional Love Songs Archiving Project includes audio recordings and an annotated Toolbox text. The Toolbox (text) file contains transcriptions for all of the audio files. The metadata for each audio file includes a reference to a specific id in the Toolbox file.

平江传统灯戏调保存项目包括录音文件和一个Toolbox文本文件。Toolbox文本文件包含所有录音文件的歌词。每个录音文件的metadata有一个对应号码，跟Toolbox文本文件的号码相对应。

Your access

 Your roles:

Depositor

Shenkai Zhang


 Nationality: Chinese
Affiliation: ILCAA

Tools

[Download metadata](#)




PARADISEC

Catalog

Home

Collections

Items

Contact

Welcome to the catalog of the **PARADISEC** collection



“Reach” 3

- *discovery*: drawing on understandings of audiences in order to help them browse, navigate, search, identify and select their items of interest

- Talk (62)
- Chatting (9)
- Sickness (9)
- Bird Story (7)
- Where Wild Things Are (5)
- Coconut Oil (3)
- Cardinal (2)
- Devilish Pig (2)
- Flying Fox and Parro
- Laplap (2)
- Linguo-labials (2)
- Prawn (2)
- Swadesh (2)
- Turtle and Shark (2)
- Ais Island (1)
- Akalao Bird and Dau
- Akalao and Mother (
- Aore Island (1)
- Before Going to War
- Circumcision (1)
- Conch and Sea Snail
- Directions (1)
- Dying (1)
- Engagement (1)
- First Coconut (1)
- Five Fingers (1)
- Numbers (1)
- Pig Attack (1)
- Pig-killing Ceremony
- Piria (1)
- Pledge (1)
- Plover and Red-head
- Rat, Short-leg and O
- Six Sisters (1)
- Surae (1)
- Troll (1)
- Turtle and Old Man (1)
- Tutuba Wild Man (1)
- Two Wild Men (1)
- Wedding (1)
- White Heron (1)
- Wild Apple (1)

Find res

Reset

Access

U R C

U R C

U R C

Language

Mavea

Bislam

English

Type

Audio (

Transc

Video (

Image

Spread

Genre

Conver

Narrati

Person

Chroni

Kaston

Proced

Descrip

Word I

Person

Speech

Metada

Song (

Topic

Talk (6

Chattir

Sickne

007mavea

Not logged in. [Login](#) | [New user](#) | [Feedback](#) | [ELAR catalogue](#)

Documentation of Mavea

[Home](#) [Metadata](#) [Discussion](#) [Resources](#)

Contributor

- Valerie Guerin (162)
- Gabriel Torno (2)

Participants

- Sera Lima Lowet (88)
- Allan Natu Lowet (18)
- Elsie Fopua Kaman (15)
- Mosela Vomei Kaman (10)
- Valerie Guerin (5)
- Fred Kaman (4)
- Jo Tavon Livo (4)
- PFL (4)
- Paul Sope Livo (4)
- James Sesei Livo (3)
- Lowet Daldal Morris (3)
- Pupu Moldovo Morris (3)
- Rogen Molavea Lowet (3)
- Gabriel Torno (2)
- John Molsi Livo (2)
- Judy Vokarae Livo (2)
- Morris Tov'aoi Kaman (2)
- Peter Vuropaitia Lowet (2)
- Alfred Moltas Kaman (1)
- Johnatan Nono Livo (1)
- Jona Parparu Morris (1)
- Lina Vatari Simptia (1)
- Rolin Vofti (1)



between June 2005 and December

ted by Gabriel Torno, entitled 'The
ght for a hat. The deposit
ers narrating this story: to listen,
d panel.

many of which come with time-

sonal histories.

hey are represented in the

as p, v, and m.

Depositor

Valérie Guérin

Nationality: French**Affiliation:** University of Hawai'i
Manoa

Your access

Default access protocol: **U R C S**Your access roles: **U R C S**

Deposit

Group represented: Mavea
speaking community, located on
Mavea Island, and Deproma,
Espiritu Santo Island, Vanuatu.

Location: Recordings were all
created in Vanuatu. Locations
include: Vunopuma, Saoroi, and
Vunopua (on Mavea Island),
Deproma (Espiritu Santo Island),
Port Vila (Efate Island), Aore and
Tanna Islands.



“Reach” 4

- *delivery*: making available selected resources according to users' preferences whether by download, view-in-browser, through apps or other means; also considering emerging devices

Choguita Rarámuri description and documentation

Search this deposit

[Reset keywords](#)

Access protocol

 (1)

Language

Choguita Rarámuri (1)

Type

Video (1)

Genre

 Conversation (1)
 Procedure (1)

Participants

MDH (1)

[Home](#) [Resources](#) [Comments](#)

Found 1 bundle in this deposit (page 1 of 1)

▼ [batusía chukú](#)

batusia_chuku.mpg

 Access protocol:

[download](#)

ID: batusia_chuku
 Title: batusía chukú
 Date created: 10/9/07
 Description: Video recording of MDH preparing the dough for tortillas

 Keywords: [Choguita Rarámuri](#) - [Conversation](#) - [Procedure](#) - [MDH](#)

Depositor

Gabriela Caballero



Nationality:

Affiliation: University of California, San Diego

Access

Your roles:

Deposit

Group represented: Rarámuri from Choguita, Guachochi municipality

Location: Mexico, Chihuahua, Guachochi municipality, Choguita




“Reach” 5

- *access management:*
 - resource delivery follows depositors’ and communities’ preferences
 - access conditions and actions are transparent
 - users have ways of applying for and negotiating for access (where possible)

Documentation of Kubokota

This deposit has been accessioned.

Search this deposit

[Reset keywords](#)

Access protocol

- URCS (155)
- URCS (3)

Language

- Kubokota (148)
- Luqa (10)

Type

- Audio (151)
- Document (4)
- ELAN (98)
- Image (11)
- Text (93)

Genre [more](#)

- Custom description (13)
- Custom narrative (16)
- Custom story (1)
- Directional story (4)
- Elicitation (17)
- [more ...](#)

Topic [more](#)

- Abandonment (1)
- Adultery (1)
- Artefacts (1)

Access protocol

- URCS (155)
- URCS (3)

Summary

This deposit contains audio files of a language spoken on Ranongga Island, Solomon Islands, resulting from fieldwork conducted between September 2006 and June 2007.

Group represented

Kubokota, Ranongga Island, Solomon Islands

Language information

Kubokota (also known as Ganoqa or Ghanongga); Luqa (Lungga)

Deposit contents

The deposit comprises over 150 audio files, as well as written texts, elicitation materials and participant observation notes.

Genres include traditional narratives, procedural and route descriptions, personal stories and accounts of everyday events in which the researcher participated. Elicited materials include responses to the caused positions and



Documentation of Kubokota

Home Resources Comments

This deposit has been accessioned.

Search this deposit

Search

Reset keywords

Access protocol

URCS x
URCS

Language

Kubokota (3)
Luqa

Type

Audio (2)
ELAN (2)
Text (1)
Document
Image

Genre [more](#)

Custom description (1)
Custom narrative (1)
Custom story
Directional story
Elicitation
[more...](#)

Topic [more](#)

Origin (1)
Ragomo (1)
Spirit (1)

Found 3 bundles in this deposit with keyword **URCS x** (page 1 of 1)

► **Obobulu people's origins**

Custom narrative
Recorded on 23/11/2006
Where the people of Obobulu came from

Keywords: Kubokota - Custom narr

▼ **Spirit**

a015IL.eaf

This file is not available to all users. If you must apply to the depositor for access, please email [chambers@elar.ac.uk](#) with an email with the outcome of your request. If approved, you still may not be allowed to view this particular file. [Apply for access](#)

a015IL.txt

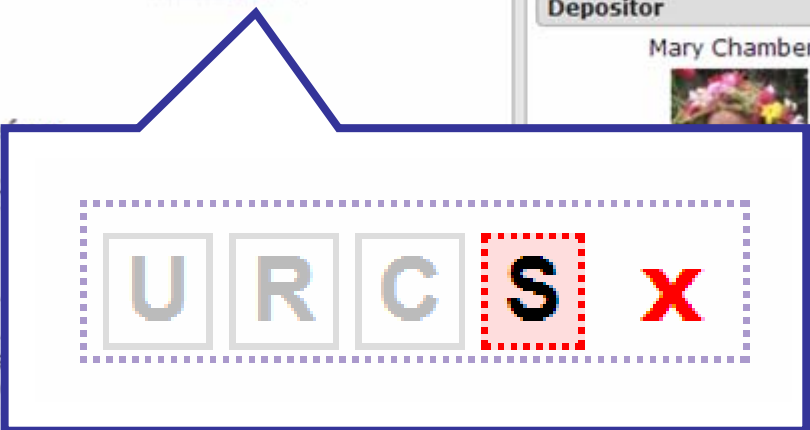
ID: a015IL
Title: Spirit
Alternative title: Ragomo
Date created: 12/11/2006
Location: Obobulu
Description: Ilikera's e... including objects a... is not go... inside the... ragomo under the washing lines where women's

Your access

Your roles: URCS

Depositor

Mary Chambers



order to be considered for access, you must apply for access rights. You will receive an email that even if your request is approved, you may not be allowed to view this particular file. [Apply for access](#)



Documentation of Ratahan

[Home](#) [Resources](#) [Comments](#)

This deposit has been accessioned.

Search this deposit

Reset keywords

Access protocol

U R C S x
U R C S (1)

Type

Audio (1)
ELAN (1)
Video (1)
Document
Text

Participants

TT (1)
WK (1)
??
???
AJ
[more ...](#)

Found 1 bundle in this deposit with keyword **U R C S x** (page 1 of 1)

▼ [RTH067 weli](#)

Access

Your roles: **U R C S**

Depositor

Anthony Jukes

Apply for access

Concerning Edward Garrett's application

Tell the depositor who you are and why you would like access to restricted material in the deposit.

Hi Anthony, I'd like to hear the recording of the 'weli' conversation



00:07 11:07

[download](#)



User

AILLA: Access

www.ailla.utexas.org/site/access_restrict.html

DOBES ELAR DELAMAN OLAC Language Res... AILLA UNESCO Speech bubble Timan

Browse

Depositors

- Formats
- Access
- Metadata
- Delivery

Rights

Languages

Events

AILLA papers

Contact

Help

Links

Access levels

We recommend that you make publicly accessible any materials that are not sensitive by their nature; that is, that are not sacred or dangerous or embarrassing for the speakers. This is best over the long term for speakers and researchers alike. If you wish to discuss your particular circumstances, please [contact us](#).

Level 1. Public access

Users have full access to these materials after agreeing to our Terms and Conditions and logging in. Choose this level for all non-sensitive materials. This is the default level.

Level 2. Password

You define a password with an optional hint. You can provide access to others by sharing the password with them. Passwords with hints allow a set of people to share resources easily. Hints can be something like that everyone in your in-group knows, but no outsider would know. Example: password = 7awin7; hint = 'zoque amigo'. You can use one password + hint for one set of resources and another for a different set. This is the most flexible option.

Level 3. Time limit

After the date specified, access will change to Level 1. Example: 2050-01-01 (January 1, 2050.) This option is good for resources that are only sensitive for a period of time; for example, the lifetime of the narrator or the five years until you finish your thesis. You will still be able to access these materials yourself with the password.

Level 4. Depositor control

Users must contact you directly to ask you for the password. AILLA will provide contact information. Note that you must maintain your contact information and be prepared to respond to requests from people you do not know.

“Reach” 6

- *information accessibility*: the actual desired content is accessible to users, whether in terms of contextualisation or appropriate complexity, language, or modality



This (lady) is the mother of my grandmother

هاي أم-ال-سني



It was she, she was considered 'The mother of the House', the lady of the house.

كانت هي هي باعتبار ام ال حوش شي ايز ذ ليدي اوف ذ هوس



In the winter

وكان بدمال-شتي



they used to make for her, a 'Kursi Jafuf'

يسوول-ها كرسي جفوف



Show more details

Click here to play/pause the whole file

Eli Timan's Jews of Iraq website: <http://www.jewsofiraq.com/texts/kersijafuf001.xml>

html 5 scroller

lah.soas.ac.uk/projects/dev/bubble-player/wilbur.html

Try repositioning and resizing the bubbles. Click on the bubble to cycle through different bubble styles. Then click on the button to show the new CSS settings.

jå jå suavastimä

bruhkujdä dä suavastit vuastav aj

text ftrE ftrS

ftrLing

Speech bubble demo: <http://lah.soas.ac.uk/projects/dev/bubble-player/wilbur.html>

“Reach” 7

- *promotion*: depositors, funders, archives and others can promote awareness
 - Wilbur (LDD12): promotion through local archiving
 - Johnston & Schembri: promotion through teaching and social media - 95 ELAR users, of which 2/3 registered in last 12 months

Media | DOBES

dobes.mpi.nl/media/?id=1839/00-0000-0000-0018-5DA7-5&type=4&project=morehead

DOBES ELAR DELAMAN OLAC Language Res... AILLA UNESCO Speech bubble Timan



DOBES

DOCUMENTATION OF ENDANGERED LANGUAGES

MOREHEAD (NEN/KÓMNZO) MEDIA

MOREHEAD (NEN/KÓMNZO) / TCI20100901-01CDV-03.MP4 (VIDEO)

Type your search term here
Search the Morehead (Nen/

Access and Registration
Research Portal
General Interest Portal
Deposit your Data
Documentation Projects
Research Projects
DOBES Programme
Archive Information






Session description: The Morehead culture is unique in the world in using a senary (base-six) number system. This system is traditionally employed in yam counting, which is an important ritual in the intense exchange relations between villages. These videos are from a yam counting procedure that took place in preparation of a wedding. Many unnamed participants were involved.
tci20100901-01CDv-01 = carrying the yams and preparing for counting;
tci20100901-01CDv-02 = counting; tci20100901-01CDv-03 = final count of the tarumbas
Keywords: yam-counting; procedural; ceremony; marriage (IMDI archive)



BROWSE THE MOREHEAD (NEN/KÓMNZO) CORP

“Reach” 8

- *communication ecology*: archive dissemination can be complemented by
 - other web sites (advice, activities etc)
 - Facebook pages (seeking contact, resources etc)
 - blogs
 - talks, publications
 - training
 - personal networks

Endangered Languages and Cultures

[HOME](#) [ABOUT](#) [AUTHORS](#) [COMMENTS POLICY](#) [PARADISEC WEBSITE](#)

« Workshop: Phonetics and phonology of Australian Indigenous languages

Models of community engagement: LIP discussion »

Imagine ... a world without PARADISEC

May 15, 2013, 1:21 am by Jane Simpson

Imagine ... a world without memories is the evocative and chilling title of a project organised by the National Committee of Australia for the UNESCO Memory of the World.



Memory of the World event 14/5/2013
Adelaide

ABOUT THE BLOG

Endangered Language and Cultures is a multi-authored blog about linguistics, language documentation, research technology, and generally everything to do with endangered languages and cultures. It is predominantly written by linguists Jane Simpson and Peter Austin, archivists from PARADISEC, and a whole slew of guest contributors.
More...

SUBSCRIBE

Subscribe via RSS to [entries](#) or [comments](#), or follow us on Twitter:

[Follow @paradisecblog](https://twitter.com/paradisecblog)

RECENT POSTS

19 October National Day of Mourning

Vale Peter Newton

Kenneth L. Hale Award: Claire Bowers (Yale University)

The NT Government gets animated

CALC (Central Australian Linguistic Circle) meeting

Rejoicing at AUSTRALEX

Useful and interesting websites

EL Blog

The Endangered Languages Archi



Home About ELAR links

← [Documentary Linguistics and Language Graveyards: an evening with Paul Newman](#)

[Tiddas Take Back William Dawes](#) →

 Search

The West Arnhem Land song project – podcast with Linda Barwick

Posted on [22 October, 2013](#) by [Ebany](#)



This week in the world of ELAR podcasts, [Linda Barwick](#) tells us about her joint documentation project with [Allan Marett](#) and other linguists and musicologists in West Arnhem Land, Australia. Linda talks about how the project came out of an “interest in how the musical traditions [of West Arnhem Land] relate to language and how music communities navigate around multilingual layers of

affiliation and communication.” Have a listen to the podcast at SOAS radio: [Western Arnhem Land Song Project](#).

An overview of the documentation project can be found on its [ELAR deposit page](#).

This is the blog of the [Endangered Languages Archive \(ELAR\)](#) at SOAS, University of London. ELAR preserves and publishes digital documentation of endangered languages.

Like 372

Follow @elararchive

[Subscribe via RSS](#)

Categories

Recent Posts

- [Call for participation: International Summer School in Language Documentation](#)

“Reach” 9

- *feedback channels*: users can provide feedback to depositors or enhance deposits with user-generated content; archive can provide usage information to depositors (and communities)

Approved subscribers

You can click on 'C' or 'S' to change these users' roles.

Pending subscribers

Please process these requests now

[Peter Austin](#) ([approve](#)/[reject](#))

Statistics

Files

156 Transcriber

69

167 Audio

7 Video

22 Image

1 Spreadsheet

422 Total files

Active users

128 [Kilu von Prince](#)

24 [Renée Lambert-Brétière](#)

22 [Kakia Chatsiou](#)

22 [Maria de Jesus Selene Hernandez Gomez](#)

19 [Svetoslav Minchev](#)

15 [Paul Butcher](#)

14 [Daniel Kaufman](#)

12 [Stefanie Pillai](#)

10 [Chikako Senge](#)

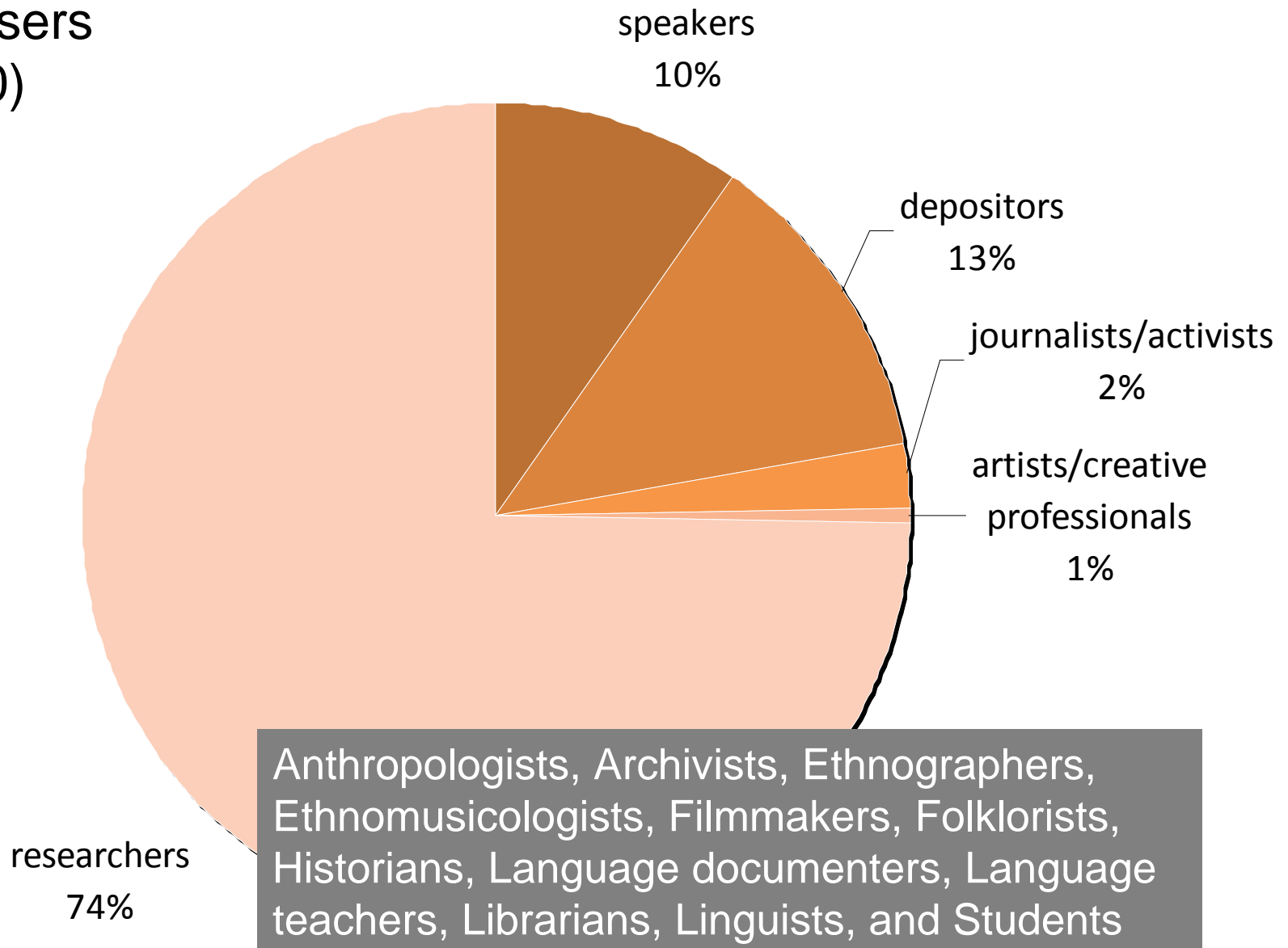
9 [Martin Haspelmath](#)

338 Total downloads

Measuring impact?

- member of OLAC and DELAMAN
- 140+ deposits, 81,000 files, 35,000 bundles, 10TB
- media (sound, video) – 47,000 recordings, 6,500 hours
- online about 700 unique visitors/day, 130,000 page views per month
- registered membership > 1100
 - > 10% community members
- but we also need to work out ways to detect and report the *value* of usage of our archives

Distribution of ELAR users (n=1040)





ELAR users tell us ...



- “I study the syntax of wh-constructions in Cape Verdean Creole and I'm interested in Portuguese contact-language situations”

“Value statements”

- “I am a teacher of the Deaf working in a Bilingual Program that uses Auslan. Use of the corpus will enable teachers and students at TPS to explore their own language and its features using real examples from a diverse range of Auslan users and range of texts”
- “I am a linguist and language teacher, I work with the Northern Arapaho language, teaching the language at the University of Wyoming, and creating instructional materials for use in all Arapaho language classrooms ”

“Value statements”

- “Kuuku Ya'u descent”
- “member of aboriginal corporation as a committee member trying to connect to my heritage through the research of our song and dance ceremonies ”
- “I am a member of the Unga tribal village and doing research for a project relating to my heritage ”
- “My son has married a goemai woman. I'm interested of this language”



Conclusion



- suggested criteria for evaluating output, impact and value of digital language archives:
 - acquisition
 - (understanding) audiences
 - discovery
 - delivery
 - access management
 - (information) accessibility
 - promotion
 - communication ecology
 - feedback

References

- Chang, Debbie. TAPS: Checklist for Responsible Archiving of Digital Language Resources (MA Thesis 2010)
- Garrett, Edward. 2013. “Participant-Driven Language archiving”. In D. Nathan & P. K. Austin (eds) *Language Documentation & Description 12*
- Wilbur, Joshua. 2013. “Archiving for the community: Engaging local archives in language documentation projects”. In D. Nathan & P. K. Austin (eds) *Language Documentation & Description 12*