

# The Role of Language Archiving in the Preservation of Local Traditions

Tom Hoogervorst ([tomhoogervorst@hotmail.com](mailto:tomhoogervorst@hotmail.com)), Leiden University  
Centre for Linguistics

*Paper presented at the*

Institute of East Asian Studies seminars (University of Malaysia Sarawak)  
14 June 2013, Kuching, Malaysia.

# Contents

- Language Archive of Insular South East Asia and West New Guinea (LAISEANG)
- Composition of the Project Team
- An area of megadiversity
- Examples of research questions
- The database
- How does this relate to Indonesia, Malaysia and East Timor?
- Lessons from the past
- Open questions
- Concluding remarks

# Language Archive of Insular South East Asia and West New Guinea (LAISEANG)

- Vision: To ensure the preservation of unique records of languages and the cultures encapsulated by them.
- Mission: To compile and archive language resources in order to create an unrivaled collection of multimedia materials and written documents from over 50 languages in Insular South East Asia and West New Guinea.
- Digitization technology:
  - Converting older recordings into WAV files
  - Scanning paper materials and converting them into PDF files
  - Metadata descriptions of language resources
  - Making the materials available online (with the exception of sensitive data)
  - Storage in The Language Archive (TLA)

# Composition of the Project Team

Dr. Marian Klamer  
Coordinator



Dr. Tom Hoogervorst  
Post-Doctoral Researcher



Chris Haskett  
Archivist, Digitization Specialist



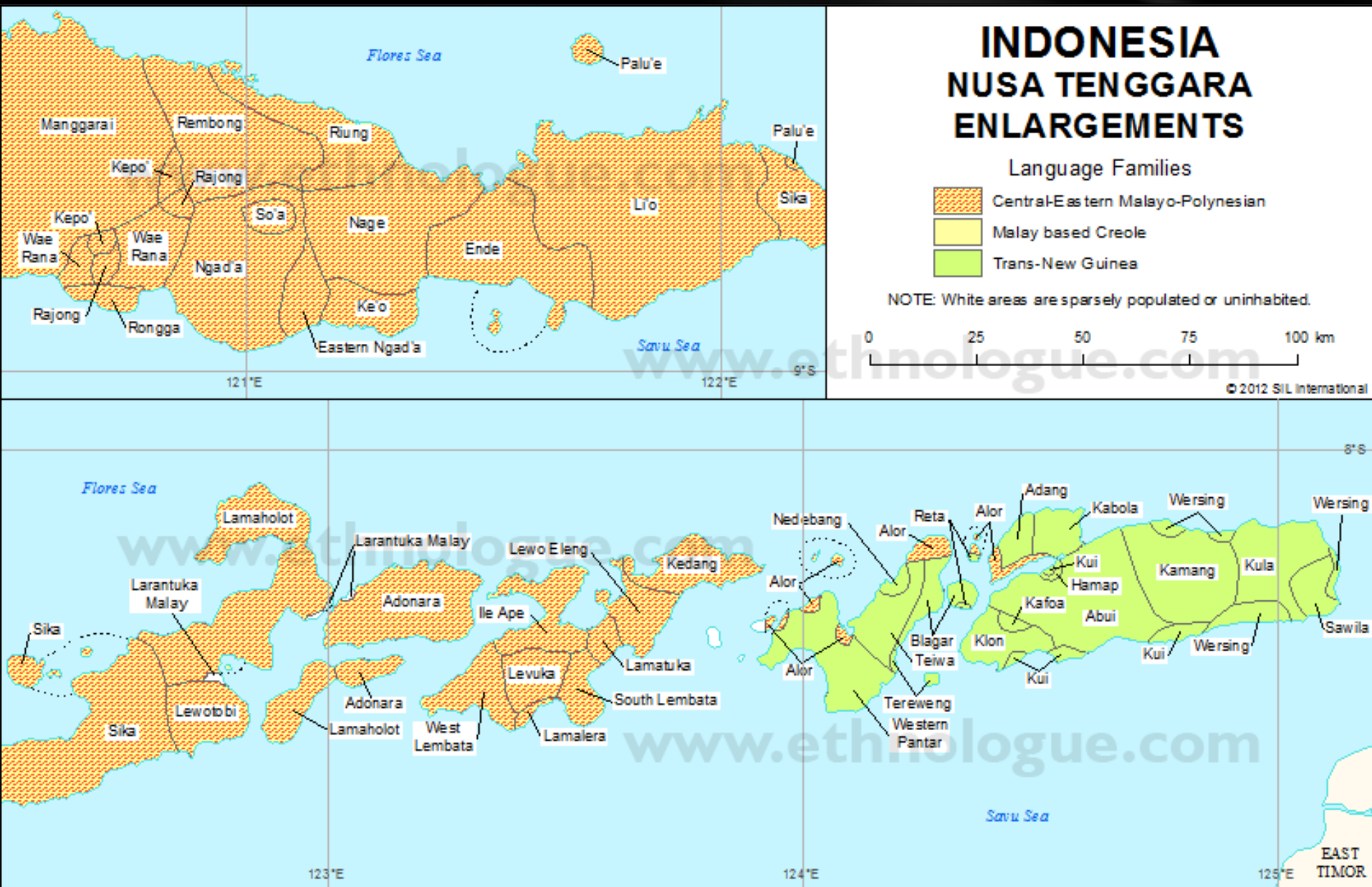
Dr. Paul Trilsbeek  
Head Archive Management



# Why digitize?

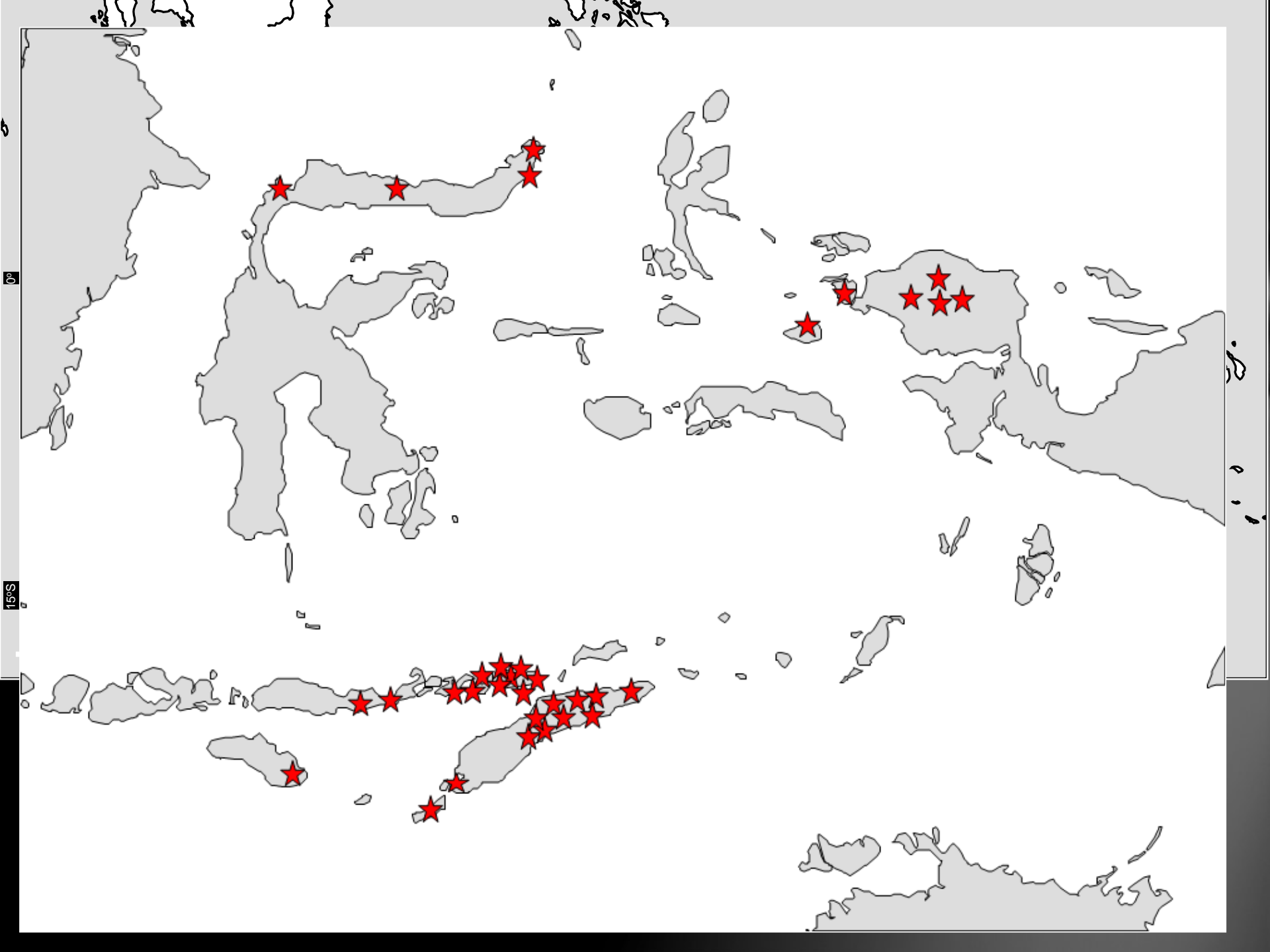
- Data is only useful if people can find and access it.
- Data about the data (metadata) –such as field notes, recordings, photographs and transcriptions– are equally important.
- The geographical region of insular South East Asia and New Guinea is well-known as an area of megadiversity.
- Both researchers and members of the speech communities should be able to use this data to answer their questions (co-operative model), even if they do not themselves know the language, and even if the language dies.

# An area of megadiversity



## Examples of research questions:

- What is the structure of language X?
- How do religious songs, historical narratives, etc. function in language X?
- Which lexical evidence supports the relatedness of different languages?
- Which numeral systems are employed in group/language X?
- What are the major intonation patterns in language X?
- What kind of omens are certain species of birds said to hold in ethnolinguistic groups X?
- What are the poetic devices and metaphors used in ritual language?
- How do/did we produce traditional material cultural items such as houses, woven cloth or baskets?





ID	Language name	Collector	Type	Content	Current media carrier	Current digital formats	Converted to LTR Standard	Annotation status	Current format of annotation	Converted to LTR Standard
1	Kambera	KLA	12 h audio	texts of various genres	audio cassettes external drive	MP3	WAV	transcribed, translated, glossed in Dutch	ascii and paper	PDF
				transcriptions of texts with morpheme breaks	external drive	ascii txt	UTF-8 txt			
		KLA	fieldnotes	transcriptions of all the texts with glosses and translations in Dutch	paper external drive	PDF	PDF	-	-	PDF
		KLA	fieldnotes	data organised according to grammatical topic (handwritten)	paper	-	PDF	-	-	PDF
		KLA		Swadesh word list with translations and references	paper	DOC PDF	PDF	-	-	PDF
2	Teiwa	KLA	7 h video	- narrative texts - structured elicitations using picture/video stimuli - traditional dances	-Sony DVM 60 Digital Video Cassette	MPEG MTS WAV	MPEG2 WAV	-transcribed -translated (English, Indonesian) -glossed (English)	Toolbox	Toolbox
3	Kaera	KLA	15 h video & audio	- texts of various genres - traditional dances - songs	- Sony DVM 60 Digital Video Cassette - mini disk	MPEG WAV	MPEG2 WAV	-transcribed -translated (Indonesian) - partly translated	Toolbox	Toolbox

50	<u>Woisika</u>	STK	25 h audio	- texts of various genres - elicitations	reels	-	WAV	partially transcribed and translated (in Dutch?)	paper	PDF
		STK	<u>fieldnotes</u>	20 notebooks of texts and grammatical elicitation	paper	-	PDF	-	-	
51	<u>Makasae</u>	HUB	1 h audio	- texts of various genres	external drive	WAV	WAV	transcribed and translated (in German?)	DOC	PDF
52	<u>Dampelas</u>	MOR	2 h audio		external drive	WAV	WAV	transcribed, glossed and translated in Indonesian	Transcriber DOC	ELAN PDF
				lexicographic materials	external drive	Toolbox	Toolbox	translated in Indonesian	Toolbox	Toolbox

Table 1. *Deliverables*

**Deliverable No**  
**Deliverable Language name**

**Due month**

1	Kambera
2	Teiwa
3	Kaera
4	Alorese
5	Alor Malay
6	Lamaholot
7	Ende
8	Roti
9	Tetun
10	Tokodede
11	Lakalei
12	Kemak
13	Tetun Dili
14	Bunaq
15	Mambai
16	Idate
17	Ambai
18	Abui
19	Sawila
20	Subo
21	Western Pantar

Project	ID	Deliverable	Date	Date completed	Comments	ARRIL ?
12-002 LAISEANG	D1	Kambera metadata + data set	5/1/13	19-2-2013	13 cables, 100mb hard drive	
12-002 LAISEANG	D2	Teiwa metadata + data set	5/1/13			
12-002 LAISEANG	D3	Kaera metadata + data set	5/1/13			
12-002 LAISEANG	D4	Alorese metadata + data set	5/1/13			
12-002 LAISEANG	D5	Alor Malay metadata + data set	5/1/13	20-2-2013		1-5-2013
12-002 LAISEANG	D6	Lamaholot metadata + data set	5/1/13	19-2-2013		
12-002 LAISEANG	D7	Ende metadata + data set	5/1/13	25-2-2013		1-5-2013
12-002 LAISEANG	D8	Roti metadata + data set	5/1/13	21-2-2013		
12-002 LAISEANG	D9	Tetun metadata + data set	5/1/13			
12-002 LAISEANG	D10	Tokodede metadata + data set	5/1/13			1-5-2013
12-002 LAISEANG	D11	Lakalei metadata + data set	5/1/13			
12-002 LAISEANG	D12	Kemak metadata + data set	5/1/13			1-5-2013
12-002 LAISEANG	D13	Tetun Dili metadata + data set	5/1/13			
12-002 LAISEANG	D14	Bunaq metadata + data set	5/1/13			1-5-2013
12-002 LAISEANG	D15	Mambai metadata + data set	5/1/13			
12-002 LAISEANG	D16	Idate metadata + data set	8/1/13			1-5-2013
12-002 LAISEANG	D17	Ambai metadata + data set	8/1/13			
12-002 LAISEANG	D18	Abui metadata + data set	8/1/13			1-5-2013
12-002 LAISEANG	D19	Sawila metadata + data set	8/1/13			
12-002 LAISEANG	D20	Subo metadata + data set	8/1/13			1-5-2013
12-002 LAISEANG	D21	Western Pantar metadata + data set	8/1/13			
12-002 LAISEANG	D22	Adang metadata + data set	5/1/13			1-5-2013
12-002 LAISEANG	D23	Inanwatan metadata + data set	5/1/13			
12-002 LAISEANG	D24	Awyu-Dumut metadata + data set	5/1/13			1-5-2013
12-002 LAISEANG	D25	Aghu metadata + data set	5/1/13			
12-002 LAISEANG	D26	Asmat metadata + data set	5/1/13			1-5-2013
12-002 LAISEANG	D27	Wambon metadata + data set	5/1/13			
12-002 LAISEANG	D28	Kombai metadata + data set	5/1/13			1-5-2013
12-002 LAISEANG	D29	Tsaukambo metadata + data set	5/1/13			
12-002 LAISEANG	D30	Citak metadata + data set	8/1/13			1-5-2013
12-002 LAISEANG	D31	Hatam metadata + data set	8/1/13			
12-002 LAISEANG	D32	Sougb metadata + data set	8/1/13			1-5-2013
12-002 LAISEANG	D33	Mansim metadata + data set	8/1/13			
12-002 LAISEANG	D34	Matbat metadata + data set	8/1/13			1-5-2013
12-002 LAISEANG	D35	Ma'ya metadata + data set	8/1/13			
12-002 LAISEANG	D36	Mpur metadata + data set	8/1/13			1-5-2013
12-002 LAISEANG	D37	Manado Malay metadata + data set	8/1/13			
12-002 LAISEANG	D38	Bantik metadata + data set	8/1/13			1-5-2013
12-002 LAISEANG	D39	Mongondow metadata + data set	8/1/13			
12-002 LAISEANG	D40	Kupang Malay metadata + data set	8/1/13			1-5-2013
12-002 LAISEANG	D41	Javanese metadata + data set	5/1/13	25-3-13		
12-002 LAISEANG	D42	Begak metadata + data set	5/1/13			1-5-2013
12-002 LAISEANG	D43	Blagar metadata + data set	5/1/13			
12-002 LAISEANG	D44	Fataluku metadata + data set	5/1/13			1-5-2013
12-002 LAISEANG	D45	Bunaq metadata + data set	8/1/13			
12-002 LAISEANG	D46	Kamang metadata + data set	8/1/13			1-5-2013
12-002 LAISEANG	D47	Abui metadata + data set	8/1/13			
12-002 LAISEANG	D48	Kemak metadata + data set	12/1/13			1-5-2013
12-002 LAISEANG	D49	Tokodede metadata + data set	12/1/13			
12-002 LAISEANG	D50	Woisika metadata + data set	12/1/13			1-5-2013
12-002 LAISEANG	D51	Makasee metadata + data set	12/1/13			
12-002 LAISEANG	D52	Dampelas metadata + data set	6/1/13			1-8-2013
12-002 LAISEANG	I-A	A synthesis of annotation issues, conventions and how they have been treated in the project, leading to a protocol/procedure for the next 6 months of the project.	6/1/13			1-8-2013
12-002 LAISEANG						1-8-2013
12-002 LAISEANG						1-8-2013
12-002 LAISEANG						1-8-2013





A27		fx KamberaTape_XIVa							
A	B	C	G	H	I	J	K	L	
1	Tape	Date	Speaker	Topic	Transcript	In electronic co	Code		
2	KamberaTape_1988_01a		1988 U Daud Maramba	Holle List (part)	In schrijfblok				
3	KamberaTape_1988_01b		1988 U Daud Maramba	Holle List (part)	In schrijfblok				
4	KamberaTape_1988_02a		1988 U Daud Maramba	(i) sentences and short story (ii) words (trans	Getypt; In schrijfblok				
5	KamberaTape_1988_02b		1988 U Daud Maramba	(i) minimal pairs (ii) Holle lists	In schrijfblok				
6	KamberaTape_1988_03a	04/04/1988	U Daud Maramba	(i) nominal phrases/compounds; Holle lists (i	Excercisebook 88-2				
7	KamberaTape_1988_03b	04-04-1988; 09-04-1988	U Daud Maramba (i, ii); U Makaborang Watuwaya (iii); G	(i) sequel to The lost Son (ii) The Poor and th	Excercisebook 88-1 (i) Verhaal A; (ii) Verhaal B; (iii) Verhaal C; (iv) Verhaa				
8	KamberaTape_1988_04a		1988 Grandfather Watuwaya; Daud Maramba; U.L. Watuwaya	(i) sequel to story (ii) reduplication (from 4 min); (iii) history of Wai Marang (from 8:19 min); (iv) minimal pairs (from 20:50					
9	KamberaTape_1988_04b		1988 U Daud Maramba	(i) sequel to translation of story grandfather (ii) reduplication old compounds (from 25 min)					
10	KamberaTape_Ja	19/02/1991	R. Miting (P); U. Tay (T); R. Danga (R); R. Hudang (H); U	(i) Conversation (ii) Religious speech with qu	Green Notebook	yes		h.v; i.v; j.v	
11	KamberaTape_Ib	19/02/1991	R. Miting (P); U. Tay (T)	Sequel of Prayer	Green Notebook	yes		j.v	
12	KamberaTape_Ila	21-02-1991; 01-03-1991	R. Paki, R. Ina, Ibu Miting, U. Li; I. MOha	(i) conversation at home (ii) watching pictures: No	No	No			
13	KamberaTape_Ilb	01/03/1991	Apu Pindi Wani; Ibu & Bpk Miting; R. Utu; R. Ana Jawa	(i) sequel watching pictures (ii) conversation \L1 utterance 1-25; L2 utterance 25-46		yes		l.v	
14	KamberaTape_Illa	13/03/1991	Ibu Miting	(i) Personal narrative about the Japanese era	Smiling Face Notebook	yes		f.v	
15	KamberaTape_Illb	(i) 06-04-1991; (ii) and ((i) R. Yana, R. Ana Hamu, Ibu & Bpk Miting (ii) and (iii) U		(i) Conversation (ii) Process description How	Smiling Face Notebook	yes		g.v; m.v; n.v	
16	KamberaTape_IVa	06/05/1991	U Prai Malinya	Narrative 'Li Nda'i' (Ancient story)	Young acress	yes		o.v	
17	KamberaTape_Va	10/06/1991	U. Yiwa Maru Matu	Narrative 'Palita Nda'i' (Old Lamp) utterance 1	Brown Notebook	yes		e.v	
18	KamberaTape_Vb	11/06/1991	U. Yiwa Maru Matu	(i) Sequel narrative 'Palita Nda'i' (Old Lamp) u	Brown Notebook; Notebook L2 (2nd half)	yes		e.v ; l.v	
19	KamberaTape_Xa	26/01/1992	U. Pindu Jawa	Sermon in church about John the Baptist (M; R		yes		r.v	
20	KamberaTape_Xb	26/01/1992	U. Hapu (H); U. Yiwa (Y); U. Raji (R); U. Pindu Jawa (P)	Meeting of congregation to discuss church at S		yes		s.v	
21	KamberaTape_Xla	16/02/1992	U. Ndena Laki Banju	Narrative Tawuru Mbiha (Magic Ring)	P1 utterance 1-58; P2 utterance 59-109; F	yes		p.v	
22	KamberaTape_Xlb	16/02/1992	U. Ndena Laki Banju	Sequel Tawura Mbiha		yes		p.v	
23	KamberaTape_Xlla	17/02/1992	U. Tai Landu Paraingu	Narrative 'Njara Hawurung' (Flying Horse)	Q1 utterance 1-59; Q2 utterance 60-114; t	yes		q.v	
24	KamberaTape_Xllb	17/02/1992	U. Tai Landu Paraingu	Sequel Njara Hawurung		yes		q.v	
25	KamberaTape_Xllla	17/02/1992	U. Musa Maramba Hau	Narrative 'Sampuraga'	T1 utterance 1-68; T2 utterance 69-96 (end)				
26	KamberaTape_Xlllb	17/02/1992	U. Musa Maramba Hau	Narrative 'Landu Niki' (Landu Niki is a name)	U entire story	yes		u.v	
27	KamberaTape_XIVa								
28									
29	Additional utterances (write Feb 1990-March 1994				Gentleman Notebook	yes		xxxx	
30									





Opnamen 1 ( -recording in vade)

	species	day	data	datum
1 RA#2-MS-01	m Hus Anfan (55)	Samate	lyt min. pare com MS	4/2/2000
2 RA#2-MS-02	f Maryam Mayallit (+30)	Samate	"	"
3 RA#2-MS-03	f Hajar Jafan (+50)	"	"	"
4 RA#2-MS-04	m Saefuddin Mayallit (25)	"	"	5/2/2000
5 RA#2-MS-05	f Ima Anfan (31)	"	1.3 min. pare	5/2/2000
6 RA#2-Banld-07468	Maes Nilah (54)	Triand	1.3 min. pare Sawdah lit dial 1,2 -> 0	11-5/2/2000 6/2/2000
7 RA#2-MS-06	m Ridwan Anfan (46)	Samate	lyt min. pare	6/2/2000
8 RA#2-MS-07	f Airun Han (35)	"	1.5 min. pare lyt min. pare	6/2/2000
9 RA#2-MS-08	m Alur Sagaf (53)	"	lyt min. pare + 3 min. pare	7/2/2000
10 RA#2-ME 01468	m Muhammad Soltip (53)	Fafanlay	Sawdah lyt 001	10/2/2000
11 RA#2-ME-02	"	"	"	dit 001
12 RA#2-ME-03	m Muhammad M Soltip (53)	"	"	dit 13/2/2000
13 RA#2-ME-04	"	"	"	dit 13/2/2000
14 RA#2-ME-05	m Ka'idat Soltip (66)	Fafanlay	lyt min. pare from ex. tran	18/2/2000
15 RA#2-ME-06	m Muhammad M Soltip (42)	"	"	19/2/2000
	kefcom ka'idat ke'west in papapara			
16 RA#2-ME-07	m Abdulgafur Ledy (50)	"	"	19/2/2000
17 RA#2-ME-08	m Muhammadita Soltip (60)	"	"	19/2/2000
18 RA#2-ME-09	m Alur Unlabet (55)	"	"	19/2/2000
	species community - slechte Devalithest			
19 RA#2-ME-10	f Halima Waynuf (29)	Fafanlay	lyt min. pare com tran, tran	20/2/2000
20 RA#2-ME-11	f Fatma Albraim (46)	Fafanlay	"	20/2/2000
21 RA#2-ME-12	f Maimuna Soltip (40)	"	"	20/2/2000
22 RA#2-Biga-1	m Christian Samigita	Biga	Sawdah dit 21/2/2000	
23 RA#2-Biga-2	m "	Biga	" dit 21/2/2000	
24 RA#2-Biga-3	m Samigita, Nad, Halilabot	Biga	min. pare dial 1.2	21/2/2000

24 + 27 + 4 = 55





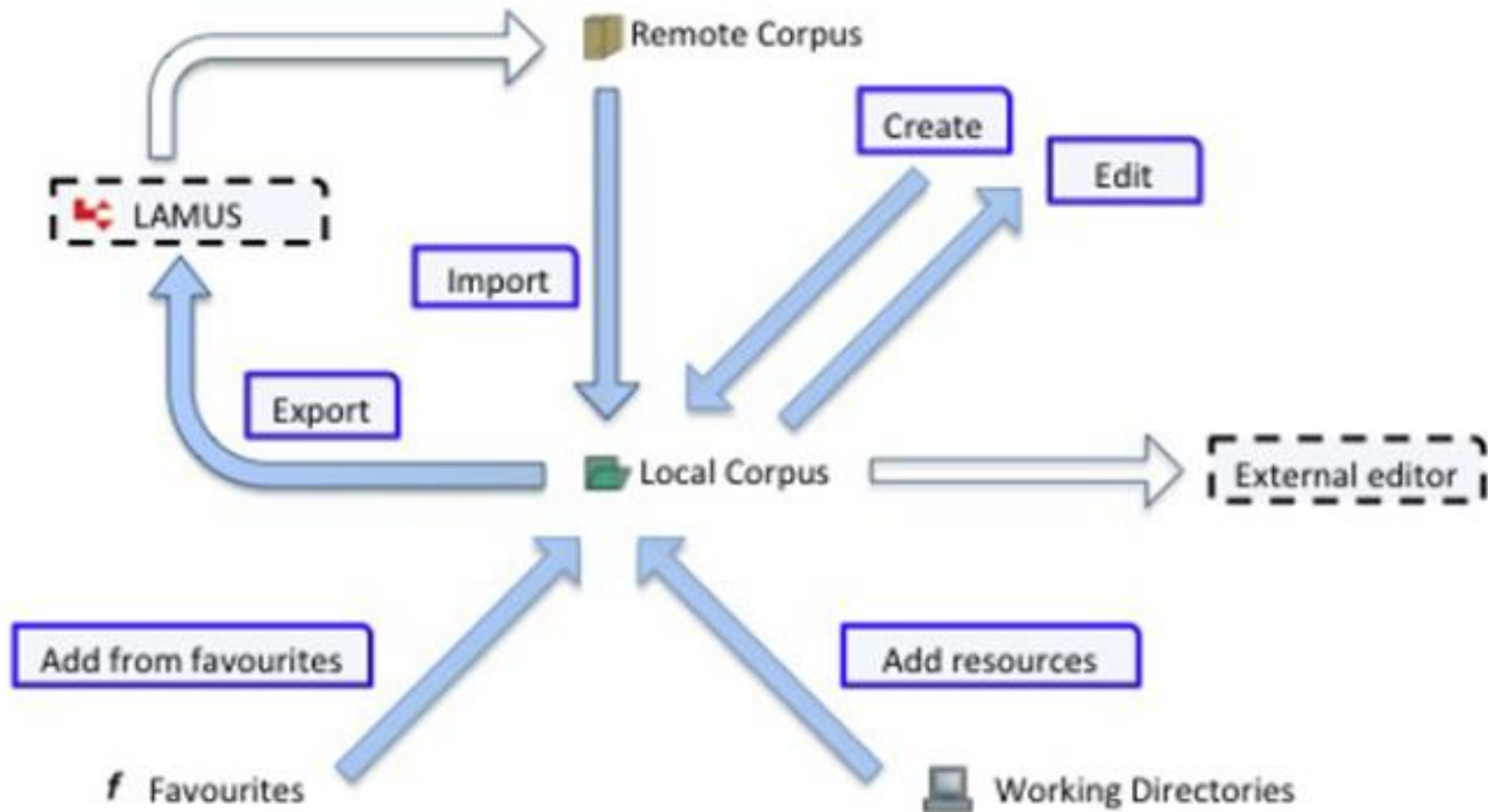




# Work Flow Overview



MAXIMILIAN-UNIVERSITÄT WÜRZBURG



# The database

Arbil Fri Feb 08 10:40:58 CET 2013

File Edit Options Window

Remote corpus

- File not found
- File not found
- File not found
- File not found
- http://corpus1.mpi.nl/d

Local corpus

- LAISEANG
  - actors (temporary node)
  - Ambel
  - Blagar
  - Butleh
  - Ende-Lio
  - Kamera
  - Lamaholot
    - Lamaholot\_I
    - Lamaholot\_II
  - Ma'ya
  - Makasae
  - Matbat
  - Roti
  - Tetun
    - Tetun\_I
    - Tetun\_II
  - Woisika

Files Favourites

Working Directories

- Alor server metadata
  - A.V
  - A.V\_Notes on NPs
  - C.V
  - C.V\_09-04-1988\_
  - D.V
  - D.V\_09-04-1988\_
  - E.V
  - E.V\_10-06-1991\_
  - F.V
  - F.V\_early 1991\_L
  - G.V
  - G.V\_06-04-1991\_
  - H.V

	Value
Collector	Hein Steinhauer
Student	Dutch, Indonesian, Russian, Croatian, French, German, English
	unspecified
	32

	Value
--	-------

3 hr 03 min (60%) remaining



# How does this relate to Indonesia, Malaysia and East Timor?

- General
  - Technological advantages
  - Demographic problems of small-scale communities



- Indonesia
  - D
  - R
  - M
- Malaysia
  - Standard
  - Culture of
- East Timor
  - The national / official language problem





**GRESIK**  
WOLU BONEK CATU, I MATEK DIAJAR SUPORTER AREMA  
WARGA HADANG SATPOL PP BOJONEGORO SAAT RA



**ORA USAH PIYA-PIYE.**  
**FAUZI BOWO PANCEN AHLINER**







## Lessons from the past

- Recordings (WAV / Mpeg-4) are extremely useful data for linguists, anthropologists, ethnomusicologists, historians and other scholars.
- For others to benefit from recordings, metadata (What? Where? When? Who?) are essential.
- To access raw data and metadata, a consistent system of archiving them will greatly help future scholars in their search through cyberspace.

## Open questions

- What is the future of the small language communities of Indonesia / Malaysia / East Timor?
- How can local communities benefit from language preservation techniques?
- What can researchers give back to the community from which they obtain their data?
- What can we learn from local notions on preservation?
- Should a language be documented when its speakers would prefer it to disappear?

## Concluding remarks

- In this digital age, the possibilities to preserve local languages and traditions are better than they have ever been in the history of mankind.
- Increased internet accessibility can connect researchers and the communities with which they collaborate. Information can and should be made available online.
- Via The Language Archive (TLA) and other documentation initiatives, the distribution of data and information has the potential to be more efficient than ever before.
- New technical developments also come with new responsibilities for the researchers: to allow others to find and access their data.

Thank you  
Terima kasih  
Matur nuwun  
Magsukul

Government of the Special Region of Yogyakarta  
Oral Traditions Association  
Universiti Malaysia Sarawak (UNIMAS)  
LAISEANG Project  
Leiden University Centre of Linguistics  
Max Planck Institute Nijmegen  
CLARIN-NL