



Online Resources on Oceanic Languages

Jacqueline Ringersma

Max Planck Institute for Psycholinguistics



In this presentation:

What is an organized, accessible and persistent archive?

Data on Oceanic languages in the MPI archive and how to access

Other online resources

Why archiving language documentation material?





Online Resources on Oceanic Languages

Language documentation:

Aim:

Maintain, consolidate or revitalize endangered languages

Through:

Creation of a representative, multipurpose and long-lasting record of languages

By:

Recording language events, speech and gesture in natural context etc.

Storing the resources in an [organized](#), [accessible](#) and [persistent archive](#)





Online Resources on Oceanic Languages



1. What's an organized, accessible and persistent archive?



What's an organized, accessible and persistent archive

Clear principles: Data organization and access infrastructure

Metadata, Online Browser and Online Metadata search (**IMDI, ARBIL**)

Coherent, consistent and persistent: Data management

Corpus structures, Archivable formats, persistent identifiers (**LAMUS**)

Safe access: Data access rights and protection

All metadata in the archive is open

All resource access can be controlled by **AMS** (web based)

Users remain the owners and stay in control of the access

Setting of licenses and code of conducts



LAT – Language archiving technology

www.mpi-lat.eu



Clear principles: Data organization and access infrastructure

What is Metadata:

Data about data

(It is the 'who, what, where and when' of a document)

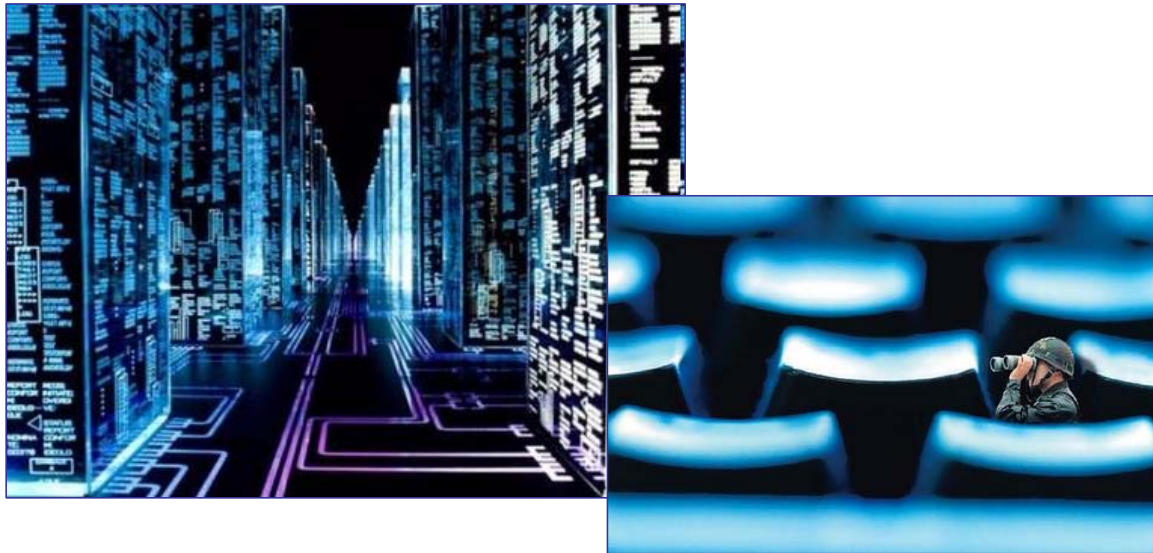
Structured data about data

Internet: machine readable data about data

Why Metadata:

Organize resources: creation of data bases according to metadata structure

(Re) Finding resources: using free text "key words" or in a structured manner



You do not want your data to float in the cyberspace and get lost!



Clear principles: Data organization and access infrastructure

Browsing the data: IMDI browser

The screenshot shows the IMDI browser interface in Mozilla Firefox. The browser address bar displays the URL http://corpus1.mpi.nl/imdi_browser/. A red box highlights the address bar, and a red arrow points to the URL <http://corpus1.mpi.nl> displayed in a white box. The main content area shows a tree view on the left and a metadata page for 'Session Iar_01' on the right.

IMDI Browser - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Back Forward Reload Stop Home

Most Visited LAMUS - Language IMDI Browser Welcome to the Max P... Tools - Language

IMDI Browser

IMDI-Browser settings manual user: anonymous login logout

IMDI ISLE Metadata Initiative

Session

Name Iar_01
Title Her work as the president of the United Church Women's Group
Date 2000-08-21

[Location](#)

[Project](#) Teop

Keys

conversion.IMDI.1.9to3.0.warning Unknown mapping of Genre: unspecified|procedure|song --> ???

[Content](#)

Actors

[Actor](#) Sao
[Actor](#) Iar
[Actor](#) Ruth Saovana Spriggs
[Actor](#)

[MediaFile](#)

[MediaFile](#)

[WrittenResource](#)

[WrittenResource](#)

[Source](#)

[Anonyms](#)

References

Tree view:

- Savosavo and Gela
- Semang
- Semoq Beri and Batek
- Sri Lanka Malay
- Taa
- Tangsa, Tai, Singpho in North East India
- Teop
 - www_mpi_nl
 - Grammar and Phonetics
 - ojs_statsbiblioteket_dk
 - 1 Teop Sketch Grammar
 - 2 Phonetik des Teop
 - 3 Articles on Teop Grammar
 - Lexicographic work
 - Teop Language Corpus
 - Culture
 - Ceremonies
 - Cultural practices and concepts (closed)
 - Cultural practices and concepts (open)
 - Material culture
 - Women (closed)
 - Iar_01
 - Iar_01.m4a
 - Iar_01.wav
 - Iar_01.eaf
 - Iar_01_RSR_teop.pdf
 - Nan_01
 - Sapia
 - Vosunana1
 - Women (open)
 - History
 - Literature
 - Phonetics
 - Vocal music
 - Tirma
 - Tofa team
 - Totoli
 - Trumai
 - Tsafiki project
 - Uru-Chipaya languages
 - Vurès and Vera'a
 - Waimaa team
 - West Ambrym Languages
 - Wichita
 - Yurakaré
 - Akhoe Hailom
- ECHO
- ESF corpus
- Endangered Languages
- GTRP Corpus



Clear principles: Data organization and access infrastructure

Browsing the data: IMDI browser

IMDI Browser - Mozilla Firefox

http://corpus1.mpi.nl/ds/imdi_browser/

IMDI Browser

settings manual user: anonymous login logout

Tree browsing

- Carvava and Gela
- Semang
- Semoq Beri and Batek
- Sri Lanka Malay
- Taa
- Tangsa, Tai, Singpho in North East India
- Teop
 - www_mpi_nl
 - Grammar and Phonetics
 - ojs_statsbiblioteket_dk
 - 1 Teop Sketch Grammar
 - 2 Phonetik des Teop
 - 3 Articles on Teop Grammar
 - Lexicographic work
 - Teop Language Corpus
 - Culture
 - Ceremonies
 - Cultural practices and concepts (closed)
 - Cultural practices and concepts (open)
 - Material culture
 - Women (closed)
 - Iar_01
 - Iar_01.m4a
 - Iar_01.wav
 - Iar_01.eaf
 - Iar_01_RSR_teop.pdf
 - Nan_01
 - Sapia
 - Vosunana1
 - Women (open)
 - History
 - Literature
 - Phonetics
 - Vocal music

- Tima
- Tofa team
- Totoli
- Trumai
- Tsafiki project
- Uru-Chipaya languages
- Vurès and Vera'a
- Waimaa team
- West Ambrym Languages
- Wichita
- Yurakaré
- #Akhoe Hailom
- ECHO
- ESF corpus
- Endangered Languages
- GTRP Corpus

ISLE Metadata Initiative

Name Iar_01
Title Her work as the president of the United Church Women's Group
Date 2000-08-21

Location
Continent Oceania
Country Papua New Guinea
Address Niu Lendi

Project Teop
Name Teop
Title Documentation of the Teop language
Id TE

Contact Seminar fuer Allgemeine und Vergleichende Sprachwissenschaft (SAVS)
Name Seminar fuer Allgemeine und Vergleichende Sprachwissenschaft (SAVS)
Address Ohlshausenstr. 40, 24098 Kiel
Email admin@linguistik.uni-kiel.de
Organisation University of Kiel

Keys
conversion.IMDI.1.9to3.0.warning Unknown mapping of Genre: unspecified|procedure|song -> ???

Content

Actors

Actor Sao
Role interviewer
Name Sao
FullName Sao
Code Sao
FamilySocialRole Unspecified

Languages
Language English
Language Tok Pisin
Language Teop

EthnicGroup Teop
Age Unknown
BirthDate Unspecified
Sex Female
Education unknown
Anonymized false



Clear principles: Data organization and access infrastructure

Browsing the data: IMDI browser

IMDI Browser - Mozilla Firefox

http://corpus1.mpi.nl/ds/imdi_browser/

IMDI Browser

IMDI-Explorer

Savosavo and Gela
Semang
Semoq Beri and Batek
Sri Lanka Malay
Taa
Tangsa, Tai, Singpho in North East India
Teop
www_mpi_nl
Grammar and Phonetics
ojs_statsbiblioteket_dk
1 Teop Sketch Grammar
2 Phonetik des Teop
3 Articles on Teop Grammar
Lexicographic work
Teop Language Corpus
Culture
Ceremonies
Cultural practices and concepts (closed)
Cultural practices and concepts (open)
Material culture
Women (closed)
Iar_01
Iar_01.m4a
Iar_01.wav
Iar_01.eaf
Iar_01_RSR_teop.pdf
Nan_01
Sapia
Vosunana1
Women (open)
History
Literature
Phonetics
Vocal music
Tima
Tofa team
Totoli
Trumai
Tsafiki project
Uru-Chipaya languages
Vurès and Vera'a
Waimaa team
West Ambrym Languages
Wichita
Yurakaré
#Akhoe Hailom
ECHO
ESF corpus
Endangered Languages
GTRP Corpus

IMDI ISLE Metadata Initiative

Session

Name Iar_01
Title Her work as the president of the United Church Women's Group
Date 2000-08-21

Location

Continent Oceania
Country Papua New Guinea
Address Niu Lendi

Project Teop

Name Teop
Title Documentation of the Teop language
Id TE

Contact Seminar fuer Allgemeine und Vergleichende Sprachwissenschaft (SAVS)

Name Seminar fuer Allgemeine und Vergleichende Sprachwissenschaft (SAVS)
Address Ohlshausenstr. 40, 24098 Kiel
Email admin@linguistik.uni-kiel.de
Organisation University of Kiel

Keys

conversion.IMDI.1.9to3.0.warning Unknown mapping of Genre: unspecified|procedure|song -> ???

Content

Actors

Actor Sao

Role interviewer
Name Sao
FullName Sao
Code Sao
FamilySocialRole Unspecified

Languages

Language English
Language Tok Pisin
Language Teop

EthnicGroup Teop
Age Unknown
BirthDate Unspecified
Sex Female
Education unknown
Anonymized false

Metadata



Clear principles: Data organization and access infrastructure

Browsing the data: Metadata search

The screenshot shows the IMDI Browser interface in Mozilla Firefox. The browser window title is "IMDI Browser - Mozilla Firefox". The address bar shows the URL "http://corpus1.mpi.nl/ds/imdi_browser/". The browser's address bar and tabs are visible, including "Most Visited", "LAMUS - Language Ar...", "IMDI Browser", "Welcome to the Max P...", "Tools - Language Arc...", "Annex embedding - ...", "Quick Store and View ...", and "LAT News".

The main content area is titled "IMDI Browser" and features a search interface. The search criteria are:

- Standard search Show occurrences
- within 1 selected corpus: DoBeS archive [13983 sessions/catalogues]
- Session.Content.Languages.Language.Name: [Dropdown menu]
- Session.Location.Country: Papua New Guinea [Dropdown menu]
- Session.Content.Genre: Discourse [Dropdown menu]

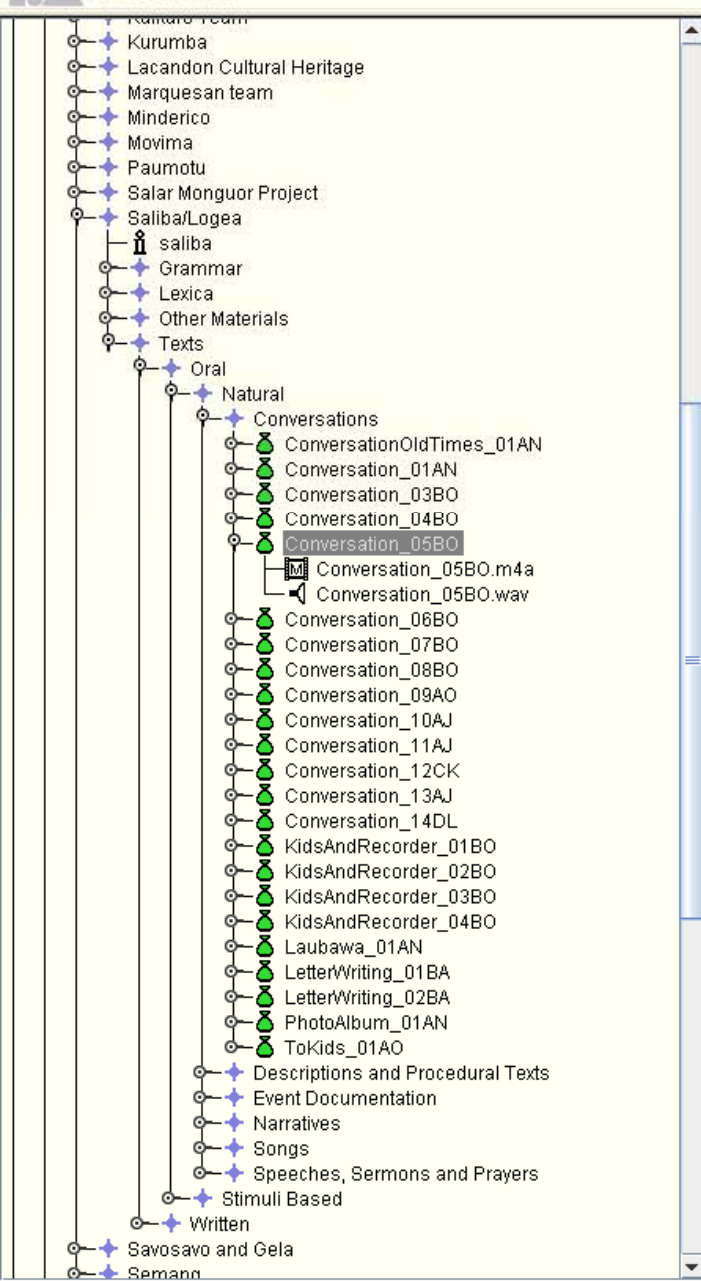
Buttons for "Search", "Save", "Print", and "Content search" are located below the search criteria. A "Help" button and "Key word search" / "Advanced search" buttons are also present.


The search results section displays "388 matches within 13983 selected sessions/catalogues found." and "Search complete". The results list includes various session paths, such as:

- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/ConversationOldTimes_01AN
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/Conversation_01AN
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/Conversation_03BO
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/Conversation_04BO
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/Conversation_05BO
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/Conversation_06BO
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/Conversation_07BO
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/Conversation_08BO
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/Conversation_09BO
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/Conversation_10BO
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/Conversation_11AJ
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/Conversation_12CK
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/Conversation_13AJ
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/Conversation_14DL
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/KidsAndRecorder_01BO
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/KidsAndRecorder_02BO
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/KidsAndRecorder_03BO
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/KidsAndRecorder_04BO
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/Laubawa_01AN
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/LetterWriting_01BA
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/LetterWriting_02BA
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/PhotoAlbum_01AN
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Conversations/ToKids_01AO
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Descriptions and Procedural Texts/AboutDialects_01DP
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Descriptions and Procedural Texts/AboutDialects_02DS
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Descriptions and Procedural Texts/Activities_01DP
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Descriptions and Procedural Texts/Activities_02CM
- /IMDI-corpora/Endangered Languages/DoBeS archive/Saliba/Logea/Texts/Oral/Natural/Descriptions and Procedural Texts/Adoption_01AO

A context menu is visible over the search results, containing the following options:

- Show Session with Tree
- Show Session
- Content search
- Show/Add Webnote(s)




ISLE Metadata Initiative

Session

Name Conversation_05BO
Title Conversation 05
Date 1996-06-06

Description
 Conversation 05

[Location](#)

Continent Oceania
Country Papua New Guinea
Region Milne Bay Province
Region Saliba
Region Sawasawaga
Address Saliba - Sawasawaga

[Project](#) Saliba/Logea

[Content](#)

Genre Discourse
SubGenre Conversation
Task Unspecified
Modalities Speech
Subject Unspecified
Interactivity non-interactive
PlanningType spontaneous
Involvement Unspecified
SocialContext Family
EventStructure Conversation
Channel Face to Face

Languages

[Language](#) Saliba

Description
 Conversation between several speakers

Actors

[Actor](#) BO

[Actor](#) Anna Margetts

[MediaFile](#)

[MediaFile](#)

[Source](#)



Clear principles: Data organization and access infrastructure

Browsing the data: resources

The screenshot shows a Mozilla Firefox browser window with the URL `http://corpus1.mpi.nl/ds/imdi_browser/`. The browser's address bar and menu bar are visible. The main content area is titled "IMDI-Browser" and displays metadata for a session. The left sidebar shows a tree view of language resources, with "lar_01" highlighted in a red box. The main content area displays the following metadata:

IMDI ISLE Metadata Initiative

Session

Name Iar_01
Title Her work as the president of the United Church Women's Group
Date 2000-08-21

Location

Continent Oceania
Country Papua New Guinea
Address Niu Lendi

Project Teop

Name Teop
Title Documentation of the Teop language
Id TE

Contact Seminar fuer Allgemeine und Vergleichende Sprachwissenschaft (SAVS)

Name Seminar fuer Allgemeine und Vergleichende Sprachwissenschaft (SAVS)
Address Ohlshausenstr. 40, 24098 Kiel
Email admin@linguistik.uni-kiel.de
Organisation University of Kiel

Keys

conversion.IMDI.1.9to3.0.warning Unknown mapping of Genre: unspecified|procedure|song -> ???

Content

Role interviewer
Name Sao
FullName Sao
Code Sao
FamilySocialRole Unspecified

Languages

Language English
Language Tok Pisin
Language Teop

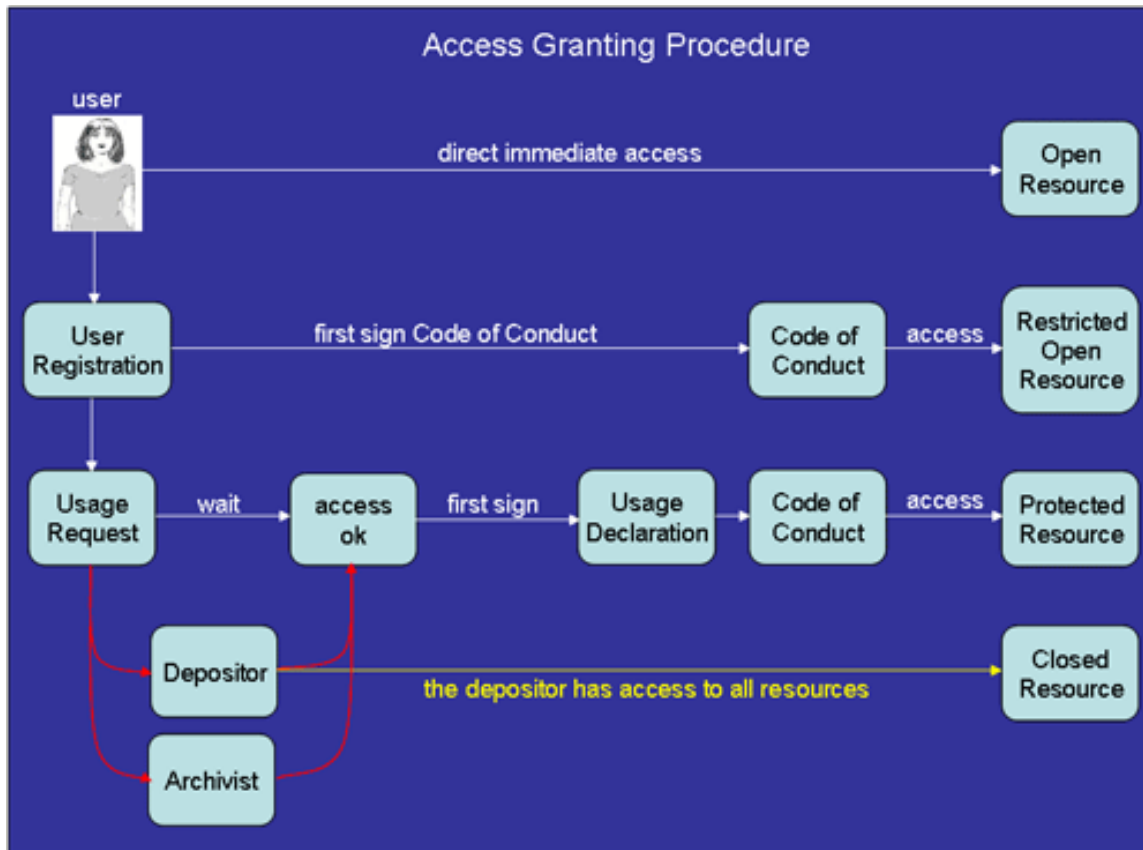
EthnicGroup Teop
Age Unknown
BirthDate Unspecified
Sex Female
Education unknown
Anonymized false

Resources



AMS on resources in IMDI browser

Access on resources: IPR, privacy, copyright etc.
(Metadata is always open!)



← Default setting after upload



Viewing the resources

Viewing the data: ANNEX viewer

The screenshot shows the ANNEX viewer interface within a Windows Internet Explorer browser window. The browser's address bar displays the URL: `http://corpus1.mpi.nl/ds/annex/protected/no_auth.jsp?sessionId=50AD6071C8F06AAB9DCBD239CD220F58&transferuid=1`. The browser's menu bar includes File, Edit, View, Favorites, Tools, and Help. The browser's toolbar shows the home button, search, print, and page settings options. The ANNEX interface itself has a top navigation bar with a red ribbon icon, the text "ANNEX", and links for "manual", "?", and "settings". On the right side of the top bar, it shows "user: anonymous" and a "login" link. The main content area is divided into several panels:

- Text:** A sidebar on the left with a list of options: "Text", "Grid", "Subtitle", and "Timeline". "Timeline" is currently selected and highlighted in blue.
- Video display:** A central panel containing a video player. The video shows a group of people sitting on the ground in an outdoor setting. Below the video player are playback controls (play, stop, previous, next) and a "Settings" button.
- Media information:** A panel on the right providing details about the resource:
 - Resource: demo_r03_v21_s1.eaf
 - Media file: demo_r03_v21_s1.m ...
 - Elapsed time: 00:00:15:000
 - Selected chunk:
 - Begin time: -
 - End time: -
 - Text: -
- Mini Data Frame:** A panel on the right with a large empty white space. Below it, there are controls for "Tier" (set to "none") and "Font size" (set to "14").
- Timeline:** A large panel at the bottom showing a horizontal timeline with a scale from 0:12:000 to 00:00:15:500. A red vertical line is positioned at approximately 00:00:14:500. Below the timeline, there are three horizontal bars representing different glosses:
 - Gloss2:** Contains the text "??", "stone axe business is the reason", and "and".
 - Gloss1:** Is currently empty.
 - Andrew:** Contains the text "chomo ch:aa u l:èè c'v:o ????" and "wut".

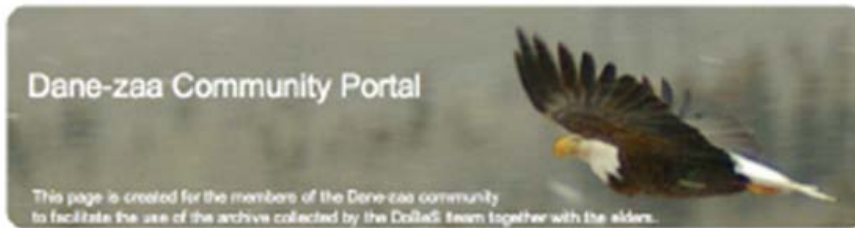
On the left side of the timeline panel, there are several navigation buttons: "Play selection", "Clear selection", and a set of directional arrows (|<, >|, <<, >>, <, >, +, -).



Viewing the resources

Different ways of accessing:

Community portals

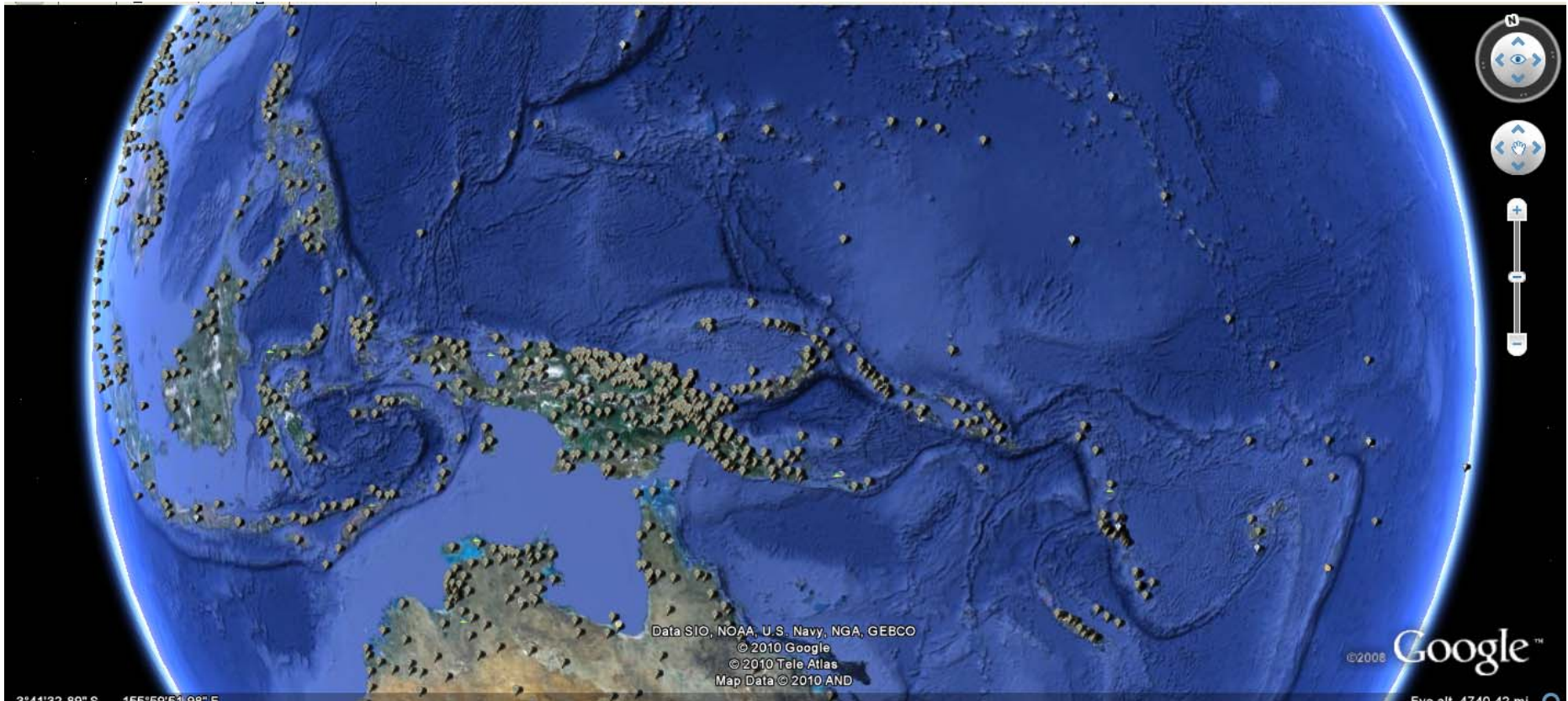


Google Earth Layers





Online Resources on Oceanic Languages



2. Data on Oceanic languages in the MPI archive and how to access



Papua New Guinea:

1. Saliba language data (text, sound, video, annotations)

The screenshot shows the IMDI Browser interface. On the left, a file tree is visible under 'Saliba-Logea', including folders for Grammar, Phonetics and Phonology, Lexica, and Texts. The main content area displays the DOBES website for the Saliba-Logea project. The website header includes the DOBES logo and navigation links. Below the header, there is a navigation menu with options like 'Language', 'Geography', 'People & Culture', 'Project', 'Team', 'Data examples', and 'Links'. A large photograph shows a traditional wooden boat on a sandy beach. Below the photo, there is a paragraph of text describing the Saliba and Logea dialects.

Saliba-Logea

Home > DoBeS Map > Documentation Projects > Saliba-Logea

Language | Geography | People & Culture | Project | Team | Data examples | Links

Documents
Help and Advice
Links
Related Events
Language Archives
Tools
LAN (Archives
Newsletter)
DELAMAN Network
DAM-LR Project
MPI for Psycholinguistics
MPI Archive

News
DOBES Training Course
June 2010
More...

Saliba and Logea are two closely related dialects belonging to the Suaucic group of the Papuan Tip Cluster of the Western Oceanic language group (Ross 1988: 190ff, Lynch et al. 2003:882). They are mutually intelligible and the main differences seem to be lexical and phonological. There has been virtually no linguistic documentation of the Logea dialect and only little on Saliba, or indeed on the Suaucic family as a whole.

English is the lingua franca of Milne Bay Province (rather than Tok Pisin as in most other parts of Papua New Guinea) and its influence is strong in all areas of daily life. English is perceived as more prestigious and is associated with better job prospects and a higher standard of living. Beyond the immediate home setting, Saliba-Logea is under great threat. The Saliba and Logea people are continuing to lead a traditional life of fishing and subsistence farming. At this point it is still possible to work with the last generation of old speakers who have essentially no knowledge of English, as well



Papua New Guinea:

1. Saliba language data (text, sound, video, annotations)

The screenshot displays the MPI ANNEX interface for accessing Saliba language data. The interface is divided into several sections:

- Top Bar:** Includes the MPI logo, navigation links (ANNEX, manual, embed, settings), and user information (user: jacrin@mpi.nl, logout).
- Left Sidebar:** Contains navigation options: Text, Grid, Subtitle, Waveform, **Timeline** (selected), and Combined.
- Video display:** Shows a video player with a play button and a progress bar. The video shows two people sitting on a mat outdoors.
- Media information:** Displays details about the selected chunk:
 - Resource: AboutDialects_01DP.i...
 - Media file: AboutDialects_01DP.i...
 - Elapsed time: 00:02:30:100
 - Selected chunk:
 - Begin time: 00:02:00:258
 - End time: 00:02:02:688
 - Text: -
- Mini Data Frame:** A large empty white area for displaying data.
- Timeline:** A detailed view of the video's content, showing a sequence of frames with their corresponding text and annotations. The text is displayed in a table-like format with columns for different parts of the video. The text includes words like "taba", "doha", "madai", "meta", "part", "n", "pro", "pref", "y", "pro", "as if", "like", "as for the 'mada'", "today we will go and do it but we haven't", "taba doha", "doha", "madai meta", "madai kabo ta lau ta ginauli na sola niga".



Papua New Guinea:

1. Saliba language data (text, sound, video, annotations, lexicon) – DoBeS team
2. Teop language (text, sound, video, annotations, lexicon) – Ulrike Mosel
3. Humanethologisches Filmarchiv – Trobiand
1979 – 1994 (Eibl-Eibesfeldt and Wulf Schiefenhövel)
4. Yélî Dnye language acquisition (video, sound, annotations) – Penny Brown (MPI)
5. Yélî Dnye language data (text, sound, video, annotations, photo, lexicon) – Steve Levinson (MPI)
6. Kilivila language data (video, sound, annotations) – Gunter Senft (MPI)
Trobriand Islanders ways of speaking book chapter links to sound and video files




Data on Oceanic languages in the MPI archive and how to access

IMDI Browser | ANNEX manual ? embed settings | user: jacrin@mpi.nl | logout

Text | Grid | Subtitle | Timeline

Video display min



Media information min

Resource: demo_r03_v20_s5.eaf
Media file: demo_r03_v20_s5.m...
Elapsed time: 00:00:19:800
Selected chunk:
Begin time: -
End time: -
Text: -

Play selection | Clear selection

Tier 1 min

Gloss1 | Gloss2 | Kp | Font size: 14 | Tier text font: Arial Unicode MS

Uou confirmation people go **You to be confirmed hurry up** This thing is talking (new version) not here, he's speaking over there he won't be speaking there He's speaking under the mangrove tree we used to start everything first That tpile we has it own start Every tpile we has it own start Yes, when I (re)start Vyuwókada flower (after midnight), then I pick up the mbwaa I won't start, this thing miught record it yhou sing it, just sing one! I also have never heard it (tricking?) I was singing it, where were you ate Kémivyuwu I was not staying close I was singing it at K I also made a mistake because of such a crowd sit down, respect this man I made a mistake, I went across Because there were a lot of people overcame me I never finished it we wanted to see you, you sing that one sing the beginning part, start at the beginning I won't sing it, people are coming closer No, just sing it differently (not loud) you go away, you Yááméné person OK, I haven't heard it, sing it in my hearing That the begginig of the VK? That's the beginning of the mbwaa section (midnight) I started doing the Myáá tpile we again because why? I might sing it now you say it properly, he might record it He's going down (to

Tier 2 min

he's speaking into his 1/2 bush knife I am saying that he's talking into a half bushknife everything we are saying is going inside See this fellow: he seems to me like a little frog it's collwecting all these things? it's recording You have sung them, sing them again You sing those two parts of the MYáá mbwaa You (P) ask him (Kp) that thing, you sung those 2 parts of Myáá, sing them again the next one I don't want to you think it will go down here (in recorder) that's where you started at the beginning of the mbwaa Ok, yu started to sing V-K flower's opening It seems to me just the same (as Myáá tpile we) sit down please I know when you started the beginning of the mbwaa the second one How many have you in memory? this one is second You sing Myáá's mbwaa first This thing isn't just anything, it's going to record it's going to record them This the timne for singing the mbwaa, it's coming out of mouth I might sing it in church You are "drinking"/singing the Tpyipwi pyu tpile we? yes, the mbwaa It's the brain of the tpile we This one gives confirmation, then he will go It's really blowing at Abaleti This guy is still here I am saying Kuungm:aa son of Kpaa at Wulé yes, I was eating taro piece and salt

Tier 3 min

?? kéle, ala yoo kamí confirm u yoo lee dmyinu **nyáá** OKay a mbwú daa k:ii, mwi wuné mbumu taa ghi u méné wuné mbumu kéle nimo taa ghi u méné wuné mbumu ?? nyáá wuné wiye mm :éé? myáá kn:áá? (touches temple) (sings verses) (sings tpile wee) mo yi mo mm ?? :ee kópu yili kn:áá yili mwiyé anyi chaa nyáá nyáá nga a ngini, ngmidi nga a ngini, ? mye doo ngé nyongo, vyuwókada néné doo ngé nyongo. nyáá :éé chono dé noo n:aa ya, kí pini? nyáá ?? nyáá ka apé! nyáá ??? kn:áá mwo anyi chaa kéle, minté a ngini ka, d:oo ngé nyongo, apé, a ngéédi nga a ngini vyuwókada néné kn:áá u chapí mbwaa kn:áá mm wu pwo ? blink-nod (blink)mm blink :aal mm Myáá p:uu médê- médê diyé cha w:ee mu kópu u léé diy:oo ka :éé :éél vyuwú kada néné mwiyé ndee :éé? (pointing at mike) nyáá mm alanté kn:áá anyi chaa, yed:oo u kada kada dpí ghé nyáá mm nyáá nyáá nyáá, mbwaa mm (nods) ?yu mbwaa daa too. nté dyéédi ye too nods (nods) nyáá u t:áát:áá a ngima mu dí kaali kí pini,dpí t:áát:áá a ngima dí kaali Y:emwe k:ii. ??mm :aa mé nté, daa proper ngé ani ghé, cha w:ee (???) ?? a koo mé kidi mé kpémikpémí, _ngoo kidi m:uu :éé nyáá nyáá nyáá, ndé kópu nyáá, kí yedé t:áát:áá ngmé ?? ló kópu

Expand | Expand | Expand

Transferring data from server...



Data on Oceanic languages in the MPI archive and how to access

French Polynesia:

1. Marquesan language data (text, sound, video, annotations, lexicon) – Gaby Cablitz

- ◆ Gente del Centro - People of the Center
- ◆ Gorani
- ◆ Hocank team
- ◆ Iwaidja team
- ◆ Jaminjungan and Eastern Ngumpin
- ◆ Kola Saami Documentation Project
- ◆ Kuikuro Team
- ◆ Kurumba
- ◆ Lacandon Cultural Heritage
- ◆ Marquesan team
 - ◆ Lexicographic data
 - ◆ Primary data
 - ◆ North Marquesas
 - ◆ Nuku Hiva
 - ◆ entretiens
 - ◆ food
 - ◆ historical sites
 - ◆ local history
 - ◆ material culture
 - ◆ narratives
 - ◆ place names
 - ◆ plant medicine
 - ◆ planting/gardening
 - ◆ poems songs
 - ◆ rites/spiritual world
 - ◆ trick language
 - ◆ Ua Huka
 - ◆ Ua Pou
 - ◆ South Marquesas
 - ◆ Fatu Iva
 - ◆ Hiva 'Oa
 - ◆ Tahuata
 - ◆ Tahiti
 - ◆ Lexicon
 - ◆ food

The screenshot displays the ANNEX media player interface. At the top, there are navigation links: ANNEX, manual, ?, embed, settings, and a user profile for Jacrin@mpi.nl with a logout option. The main content area is divided into several panels:

- Text:** A sidebar on the left with tabs for Text, Grid, Subtitle, Waveform, Timeline, and Combined.
- Video display:** A video player showing a man sitting outdoors. Below the video are playback controls: Play selection, Clear selection, and buttons for previous, next, full, and buffer.
- Media information:** A panel showing resource details: Resource: kanea_maa-TM.eaf, Media file: kanea_maa-TM.mp4, Elapsed time: 00:00:00:000, and Selected chunk information (Begin time, End time, Text).
- Mini Data Frame:** A panel for displaying data frames, currently showing 'Tier: none' and 'Font size: 14'.
- Waveform and Timeline:** A large panel at the bottom showing a waveform for the audio and a timeline with multiple subtitle tracks. The subtitle tracks include French and English text, such as 'maintenant le fruit à pain a été cueilli, il y a le fruit à pain en bas' and 'ah cette heure-ci je vais tailler la perceuse afin de percer ce machin pour hâter la matur'.



French Polynesia:

1. Marquesan language data (text, sound, video, annotations, lexicon) – Gaby Cablitz

Solomon Islands:

1. Savosavo and Gele language data (mainly audio, some video) – Claudia Wegener

Vanuatu:

1. Vera'a language data (audio, video) – Vera'a and Vures Dobes project – Stefan Schnell
2. Dakaka language data (audio, annotation) – Ambrym Dobes project - Manfred Krifka



Data on Oceanic languages in the MPI archive and how to access

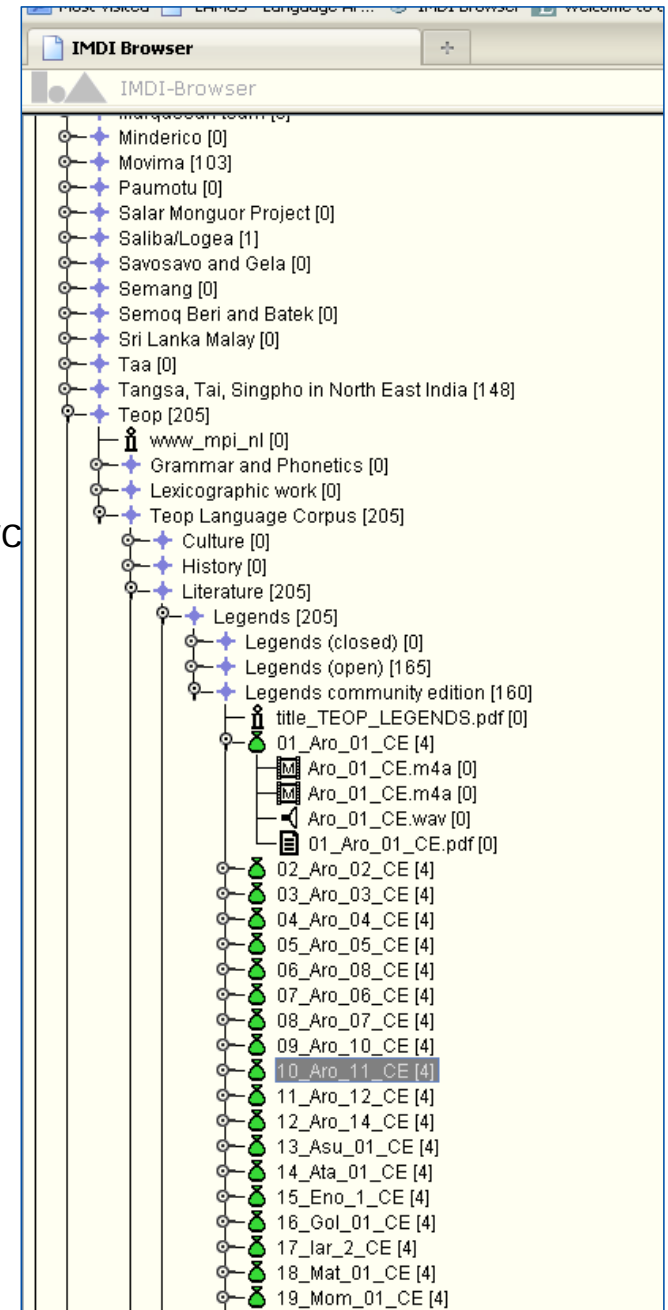
How to find the data you would like to explore?

Metadata search!

Google earth overlay (demo)

How to access the resources:

- Some are open (teasers) – Use the 'show accessible resource





Data on Oceanic languages in the MPI archive and how to access

How to find the data you would like to explore?

Metadata search!

How to access the resources:

- Some are open (teasers) – Use the 'show accessible res
- Some are closed – Use request access option.

The screenshot shows the IMDI Browser interface with a tree view of language resources. The tree structure includes:

- marquesan team [0]
- Minderico [0]
- Movima [103]
- Paumotu [0]
- Salar Monguor Project [0]
- Saliba/Logea [1]
- Savosavo and Gela [0]
- Semang [0]
- Semoq Beri and Batek [0]
- Sri Lanka Malay [0]
- Taa [0]
- Tangsa, Tai, Singpho in North East India [148]
- Teop [205]
- www_mpi_nl [0]
- Grammar and Phonetics [0]
- Lexicographic work [0]
- Teop Language Corpus [205]
- Culture [0]
- History [0]
- Literature [205]
- Legends [205]
- Legends (closed) [0]
- Legends (open) [165]
- Legends community edition [160]
- Modern Stories [0]
- Sii_04 [0]
- Sii_04.m4a [0]
- Sii_04.wav [0]
- Sii_05 [0]
- Phonetics [0]
- Vocal music [0]
- Tima [13456]
- Tofa team [0]
- Totoli [1]
- Trumai [0]
- Tsafiki project [0]
- Uru-Chipaya languages [0]
- Vurès and Vera'a [0]
- Waimaa team [2]
- West Ambrym Languages [0]
- Wichita [25]
- Yurakaré [0]
- ≠Akhoe Hai//om [1]
- ECHO

A context menu is open over the 'Sii_04.m4a' file, showing the following options:

- view node
- create bookmark
- download
- request resource access**
- set access rights
- version info



3. Other online resources



Other online resources

LEXUS & ViCoS: Web based lexicon tool, multimedia encyclopedia

The image displays two overlapping browser windows. The left window, titled 'demo rossel lexicon', shows a search interface with a 'List' tab selected. It lists search results for the letter 'm', including 'mbele' (fishing with light), 'mb:emi kóó' (fish sp. Barracuda), 'mb:emikó' (long tom), 'mbenembene' (fish sp. Flounder), 'mbimi' (bat sp. grass), 'mbipí' (fish sp. yellowfin goatfish), 'mbiye' (calling of bird), 'mbudu' (fish type green parrot fish), and 'mbuwó' (bird sp. heron). The right window, titled 'ViCoS Editor and Navigator - Mozilla Firefox', shows a conceptual network diagram. The central node is 'nmo'. It is connected to various other nodes, including 'te', 'tpile tp:oo', 'wuu', 'yomo', 'kuu', 'tii', 'ghii', 'y:emwe', 'pyipo', 'ghém', 'kidi', 'ghém', 'tá', 'kêêlvé', 'ghêpê', 'kíye p', 'kpéé', 'kiní', 'ghêêdi kuu', 'chile', 'chu', 'keepi', 'maafi', 'tada', 'taa', 'chuu', and 'mgámu'. A legend on the right side of the ViCoS window explains the relation types: 'is_father_of' (red dot), 'eats' (red dot), 'sounds' (red dot), 'is_a_kind_of' (green dot), 'is_not_a_kind_of' (red dot), 'is_part_of' (red dot), 'is_antonym_of' (red dot), 'is_synonym_of' (red dot), and 'is_related_to' (red dot). Below the legend, there are 'Modes' (browse, move, connect, delete, lexus, world, attach, detach) and 'Relation Types'.



Other online resources

LEXUS & ViCoS:

Marquesan
Tuamotuan
Yeli Dnye
Teop
Saliba

LEXUS Word List View Search Relation types Help brucebirch

Iwaidja lexicon: Word list

Results 1-25 of 3468 for all entries
<< first < previous Page 1 next > last >>

	Lexical Entry	
a	^Vira	
b	papp	
c	only	
d	-a	
f	a	
h	a-	
i	a-	
j	a-	

abangkun
"
1 big and tall
Wara warekbi abangkun.
There's a big, tall man.

http://corpus1.mp...-selectedIndex=0

File Switch to Help Admin

Saliba Lexicon: Modified MDF 4.0 type

Lexicon Filters

Switch to structure view

a <filter>

Lexical entries

- aa
- adisi
- aga
- aiyuui
- amokisa**
- anka
- ankol
- anti
- ao'ao
- apolo
- autibodi
- ayani

Lexical entry Lexical Entry View

amokisa

intrj

told.you.so/poor.one

amokisa Rini
poor Rini
amokisa!
this serves you right!



4. Why archiving?



Why archiving?

Misconceptions about archiving

1. Your stuff is buried here and gone forever



(from Andrew Garret
Berkeley Archives)



Why archiving?

Misconceptions about archiving

1. Your stuff is buried here and gone forever
2. Other linguists will take advantage of your hard work and take away your good ideas



(from Andrew Garret
Berkeley Archives)



Why archiving?

But the actual truth is that:

1. Other linguists do not really care about your work
2. The people who do care are the members of the speech communities – and they care about it in a different way than you do

“The coolest thing to do with your data will be thought of by someone else”



Why archiving?

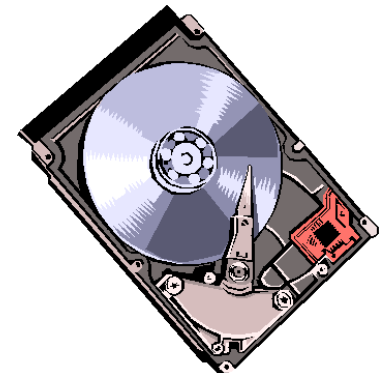
Is there a danger that we loose digital data?

YES,

UNESCO: 80% of our recordings is endangered

How much of your data and files on the notebook is organized, backed-up?

How long can media and formats be accessed?





Why archiving?

Is there a danger that we loose the data?

YES, a few messages

Archive data into a trusted archive (long term preservation and accessibility)
Create high quality metadata so that you can find the way back to the data
Use open standards

MPI-PL archive:

Trusted archive,

- ORGANIZED information
- Continuous extension
- Collaboration and interaction
- Commentary and relation drawing (enrichment)
- Supporting centre for cross-corpus and language work





Why archiving?

MPI archive is open for language data





Archiving instance, Max Planck Institute for Psycholinguistics

Archive managers: 3

Archive developers: 2

System manager: 1

Archiving software development: 4

Enrichment software development: 4

Archive for language data:

40 Terabyte of data

400.000 archived objects





Training sessions:

We regularly organize training sessions on:

Audio and video handling

Archiving technology

Enrichment of data

We do welcome participants from other than DoBeS or MPI projects

Its 4-5 days in one week

Interested: please check our MPI website/Events section (www.mpi.nl/events/)

Contact:

jacqueline.ringersma@mpi.nl

paul.trilsbeek@mpi.nl