

# Language Archiving Technology

an eHumanities infrastructure

<http://www.lat-mpi.eu>

- What is LAT?
  - History, context
  - Current state
- Case study: LAT and the corpus spoken Dutch
- Future evolutions



- 4 main research groups:
  - Language Acquisition
  - Language Comprehension
  - Language & Cognition (“field linguists”)
  - Language Production
- Large technical group: support & software development



- Started creating a digital archive about 10 years ago for a number of reasons:
  - Preventing “data cemeteries”: central, well organized data archive with rich metadata descriptions
  - Taking care of long-term preservation of data: continuous migration to latest storage technology, making use of “archivable” file formats, data distribution
  - Data access: Making data discoverable and accessible in various ways over the internet

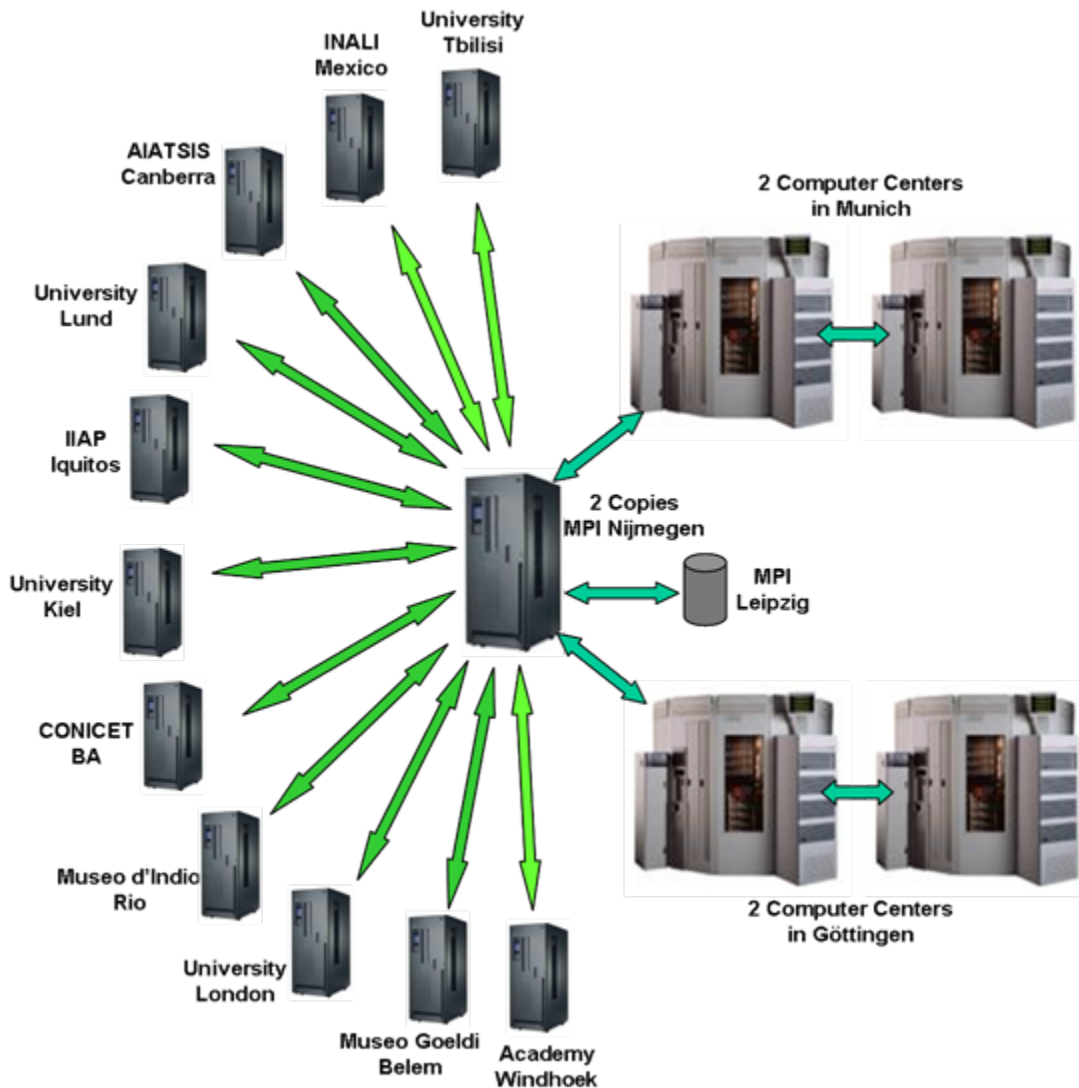


- Contains a large number of languages:
  - Languages from all over the world studied by MPI field linguists
  - First and second language acquisition studies
  - Endangered languages documented for the VolkswagenStiftung DoBeS programme
  - Spoken Dutch corpus
  - Sign languages
  - etc.



- Kinds of data:
  - video recordings
  - audio recordings
  - images
  - transcriptions/annotations
  - lexica
  - various other types of written documents (descriptions, grammars, ...)
- All described using IMDI metadata scheme

# Data distribution



- at MPI about 45 Terabyte in total (20 TB in IMDI-described corpora)
- > 250.000 resources
- > 15.000 hours recording
- every 5 years new storage technology
- 4 copies at large computer centers
- **synchronized regional archives**
  - another copy
  - data back to regions
  - trust

# Tasks of a digital archive

- long-term preservation
- giving access in the **Live** Archives sense
  - archive as a center of ORGANIZED information
  - archive as a center for continuous extension
  - archive as a center of collaboration and interaction
  - archive as a center for commentary and relation drawing (enrichment)
  - archive as a center supporting cross-corpus/language work

The image displays two related documents. On the left is a slide titled 'User Services' with the following content:

**User Services**  
Live Archives offer three main classes of user services: upload, management and access.

**Upload**  
Digital archiving is an ongoing process. Depositors and users can upload additional resources or update existing ones. Data management systems will provide facilities for upload, while tracking versions and identifiers so that important information and references will not be lost.

**Management**  
LEAs data management systems allow users to define access policies and rights, to carry out checks, to do conversions and downloads, and to view statistics about their data and the rate of access.

**Access Services**  
LEAs will offer a number of access services:

- WWW catalogue browsing and searching via metadata or via geographic interfaces, to support resource discovery
- WWW access to resources (where permissions allow) via standard web browsers
- Individual resources as well as sub-corpora, together with relevant metadata, should be downloadable to a variety of devices
- specialized applications to allow complex resources to be accessed via the WWW
- open-standard metadata (JAI-DIH, XRL-Schema) provided for harvest by other bodies to expand resource discovery potential

On the right is a brochure for 'Digital Language Resource Archives'. The title is 'Digital Language Resource Archives' and 'LIVE ARCHIVES'. It includes the following text:

Digital Language Resource Archives

**Supporting Institutions**

The intention is that as many institutions as possible that hold language resources will support the Live Archives principle and direction, as described in this booklet. If your institution may be interested in supporting it, or finding out more information, please fill in the form on the web-site:

[www.mpi.nl/dam-ly/lra-flyer](http://www.mpi.nl/dam-ly/lra-flyer)

Currently the following institutions have declared support for Live Archives (see the website for the latest information):

- ILL for Psycholinguistics, Nijmegen\*
- Centre for Languages and Literature, University of Lund\*
- ILL, Leiden\*
- SOAS, University of London\*
- Radboud University Nijmegen
- University Sheffield
- Charles University Prague
- Bulgarian Academy of Sciences, Sofia
- Institut Romantischeskoe Akademii of Sciences
- University of Bergen
- Institute of Linguistics, University of Tübingen
- University of Antwerp
- Centre de Recherches sur la Description de l'Oral, CNRS, Paris
- University of Helsinki
- Institute for Language and Speech Processing, Athens
- University of Zagreb

Other institutes that declared their support for this document are listed on the web-site.

March 2005



# MPI "LAT" Archiving Framework & Tools

## Preparation



Primary resources



ELAN/LEXUS/SYNPATHY Annotation, Lexicon & Syntax

## Integration



IMDI: Metadata descriptions, data organization



LAMUS: Data Uploading and Management, Access Management

Archive

>20 Terabytes  
>15.000 hours A/V

# MPI "LAT" Archiving Framework & Tools

## Preparation



Primary resources



**ELAN/LEXUS/SYNPATHY** Annotation, Lexicon & Syntax

## Integration



**IMDI**: Metadata descriptions, data organization



**LAMUS**: Data Uploading and Management, Access Management

Archive

>20 Terabytes  
>15.000 hours A/V

## Utilization



**IMDI/GIS**: Metadata Browsing & Searching, Google Earth



**ANNEX/LEXUS/TROVA**: Complex Access via the Web



**ADDIT/VICOS**: Enrichments/Views



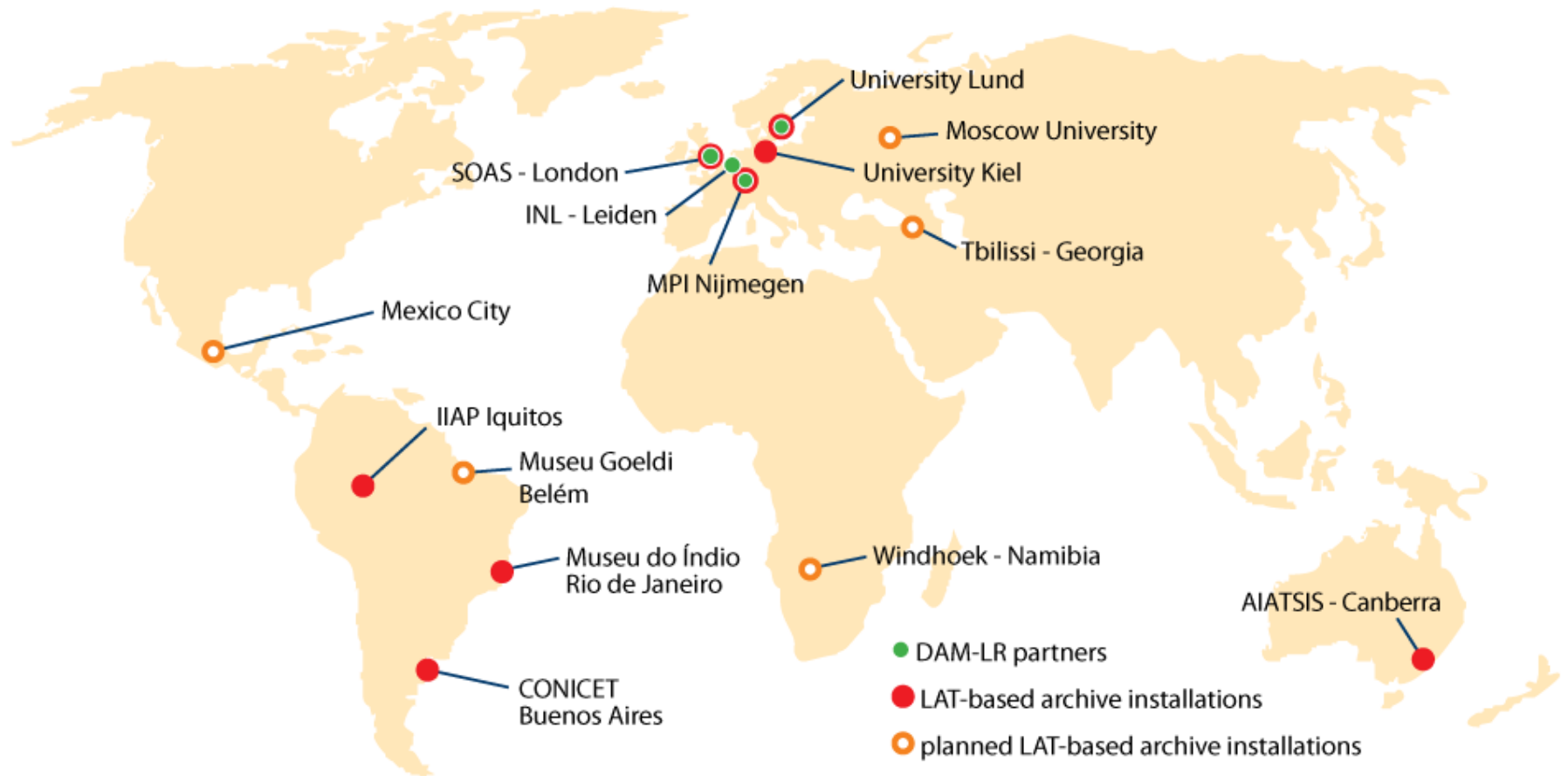
# So what is LAT about?

- Openness:
  - Open standards
  - Open source (& VMware image)
  - Long-term preservation
- Compatibility:
  - Between all LAT applications
  - With outside world (import/export)
    - Praat
    - Shoebox / toolbox / ...
    - OAI-PMH / REST-based access
- Research-centered:
  - What are researchers looking for?
  - Cross-media: links between the archive and
    - paper/ electronic publications
    - websites



# Grid of Language Archives

## Distribution of MPI “LAT” archive installations





- Example: Corpus Spoken Dutch (CGN)
  - Manually verified broad phonetic transcription: SAMPA-based
  - Prosodic annotation layer (Grover et al. 1998) :
    - strong break: ja dat weet ik || maar wanneer
    - weak break: dit is werkelijk on|ge|looflijk
    - vowel part of a prominente syllable: ^i^k ben thuis
    - unusual lengthening of a sound: %ja%



## LAT and phonological corpora (2)

- Example: find words ending in shwa, followed by a word with the stress on a syllable with /aa/



# CGN: example annotation

MPI - ANNEX Interface - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://corpus1.mpi.nl/ds/annex/protected/no\_auth.jsp?sessionid=48A1E79447DCDD283D3320C934E2012D&transferuid=1&nodeid=MPI87496%23

Most MPI watch wiki taitech clarin books tmp drupal PID dict CiteULike To Do Calendar Tel dev old ProQuest Information... drupal migration statu...

DoBeS - Training Co... IMDI\_CAICYT.pdf (ap... the Language Archivi... Workshop IMDI Browser MPI - ANNEX Interface History : fn000055 - ... MPI - ANNEX Interf... Ontwerp en motivatie

ANNEX manual ? settings user: dietuyt@mpi.nl logout

**Text**

**Grid**

**Subtitle**

**Waveform**

**Timeline**

**Combined**

**Video display** min

No streaming media available

⏪ ⏩ ⏴ ⏵

**Media information** min

Resource: fn000055.eaf  
Media file: fn000055.m4a

Elapsed time: 00:01:01:383

Selected chunk:  
Begin time: -  
End time: -  
Text: -

**Mini Data Frame** min

hartelijk d^a^nk voor uw b^i^j^drage u h^e^ft wat dat betreft als r^e^ctor naar m^i^j^n oordeel altijd een gr^o^ot v^o^o^rdeel b^o^ven de uh l^e^dinggevenden in het ^o^penbaar best^u^r want ^a^ls u d^i^t soort verpl^i^chtingen hebt zoals u vand^a^ag er m^e^r zult h^e^bben dan zijn het vr^i^j^wel altijd verplichtingen met een z^e^r inh^o^delijk wetenschappelijk kar^a^kter en dat k^a^n toch als een groot v^o^o^rrecht worden besch^o^wd dunkt mij dames en h^e^ren de ontw^i^kkelingen op het geb^i^e^d van van uh inform^a^tie- en communic^a^tietechnologie die v^o^lgen elkaar in uh in een h^e^l rap tempo ^o^p u s^o^ms heb ik wel 'ns de indruk dat de de ontw^i^kkelingen sn^e^ller gaan dan de implement^a^tie u s^o^ms heb ik daar ook wel 'ns een kl^e^ne eigen erv^a^ring bij dan h^e^b je de f^o^lders gekregen en je w^e^t wat de technolog^i^e is alleen dan bl^i^ken er geen t^e^chnici meer te zijn die dat kunnen uh implement^e^ren dat is s^o^ms zo'n kl^e^n probl^e^mpje en en uh 't kan ook wel eens een probl^e^m zijn als de polit^i^ek zich d'r ^e^igen van maakt h^e^ want IC^T^ is soms toch ^o^o^t iets van H^a^rlemmer ^o^lie

Tier: N00053-Prosody1

Font size: 14

Play selection

Clear selection

|< >|

<< >>

< >

+ -

Play screen by screen

Play continually

Tier text font: Arial Unicode MS

**Timeline**

N00053-Pos	0	ADJ(prenom,basis,zonder)	N(soort,ev,basis,zic,TSW0)	TSW0	BW0	LID(b,N(soort,ev,basis,zi,VZ(inif)	VZ(inif
N00053-Lemma		Haarlemmer	olie	hè	uh	bijvoorbeeld	de minister van van
N00053-Phonetic		harlEm@r	oli	hE	@	b@vorb@ld	d@_mnlst@r vAn vAn
N00053-Phonetic@left						SHAE	
N00053-Phonetic@right						SHARE-P(d)	
N00053-Prosody-Unit					,13		
N00053-Prosody1		H^a^rlemmer	^o^lie	hè	uh	bijv^o^rbeeld	de min^i^ster van van
N00053-Prosody1@leftb		weak			auto	strong	
N00053-Prosody1@rightb					auto	strong	

# Focus on LAT apps: ARBIL

The screenshot displays the Arbil software interface. On the left, a tree view shows a hierarchy of corpora: Remote Corpus (Beaver Archive, ECLING project, Gente del Centro - People of the Center, IMDI-corpora, MPI corpora, Acquisition, Comprehension, Demo, Ams Demo, DoBeS Demo, DOBES training, PeWi corpus, kleve-route, Actors, Languages) and Local Corpus (Buasiana2, PeWi corpus, Kleve-route, Actors, Kita, Peter, Languages, Peter Wittenburg, Peter Wittenburg, Sotaro Kita, Standard Actor, Standard Actor, Standard Actor, Standard Actor, Standard Actor, Languages, MediaFiles, WrittenResources, phidang\_talk). A 'Testing' folder contains AMS, docbook.dtd, elan-example1.eaf, elan-example1.pfs, Lamus, manualXML, oriole.eaf, and oriole.nfs.

The main area shows a table of metadata for the selected actor, Peter Wittenburg. The table has two columns: 'IMDI Field' and 'Value'. The data is as follows:

IMDI Field	Value
Role	interviewee
Name	Peter
FullName	Peter Wittenburg
Code	W
FamilySocialRole	Unspecified

Below this table is a 'Selection' window, which is currently empty. Below that is a 'Selection(1)' window, which displays the same metadata table as above, but with additional fields:

IMDI Field	Value
Role	interviewee
Name	Peter
FullName	Peter Wittenburg
Code	W
FamilySocialRole	Unspecified
Languages.Description	
EthnicGroup	
Age	Unknown
BirthDate	Unspecified
Sex	Unknown
Education	university
Anonymized	true
Contact	
Description	





# Focus on LAT apps: LAMUS

LAMUS - Language Archive Management and Upload System - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://corpus1/jkc/lamus/wsframes.jsp

LAMUS Language Archive Management and Upload System user: dietuyt logout

Sandbox

- SE-G-1
  - saludo.norma
  - saludos-imdi.eaf

### Free Media resource Nodes

linking files to **saludo.norma**

click on the table headers (twice for reverse) for sorting the rows for a certain column

Type	Name	file extension	Title	has children	view
<input checked="" type="checkbox"/> Media resource	elan-example1.wav	.wav	/home/dietuyt/doc/doc/elan_example/elan-example1.wav		view

Select All Reset All Link

Upload Files Request Storage Unlinked Files Submit Workspace Save and Logout Delete Workspace Help Report a Bug

Done

# Focus on LAT apps: Access Management

## Node Authorization Overview


- view access privileges to LAT Resources 'LAMS Demo: Lams Demo Title' -

Focus overview on selected User/Group:

### Rules of Node 'LAMS Demo'

- MPI0#/MPI1# -

Assigned Licenses:


 Webcode (webcode.html)

▶ [Manage Node Licenses](#)

 John Doe

- Read Info Files Allow
- Read Annotations Allow
- Read Images Allow
- Read Audio Files Allow
- Read Video Files Allow

▶ [Edit](#) 

 John Doe's research group

- Read Info Files Allow
- Read Annotations Allow

▶ [Edit](#) 

▶ [Add new Rules](#)

### Rules of Node 'Demo'

- MPI0# -

There are no licenses assigned to Node 'Demo'

There are no privileges directly assigned to Node 'Demo'

# Focus on LAT apps: ELAN

Elan - NGT\_AH\_fab5.eaf

File Edit Search View Options Help

CAM 3

Grid Text Subtitles Controls

Translation Dutch  
Hij keek voorzichtig rond, niemand te zien.

Translation English  
He looked around carefully, nobody there.

Gloss RH  
NIETS

Mouth  
bialabial

00:00:13.640 Selection: 00:00:13.640 - 00:00:15.650 2010

Selection Mode Loop Mode

	00:00:14.000	00:00:15.000	00:00:16.000	00:00:17.000	00:00:18.000	00:00:19.000	00:00:20.000	00:00:21.000	00:00:22.000	
Translation Dutch	emand te zien.		Hij rende snel de winkel in, pakte het bot en rende er zo snel als hij kon mee weg. Hij rende ver weg tot aan de br							
Translation English	nobody there.		He ran into the shop, took the bone and took off as fast as he could. He ran far away up to the bridge.							
Gloss RH English	NOTHING		(p-) running dog	CATCH	(p-) running d	(p-) dog disappears	BRIDGE	(p-) run		
Gloss LH English	NOTHING		(p-) running dog		(p-) running d		BRIDGE			
Gloss RH	NIETS		(p-) rennen hond	GRIJPEN	(p-) rennen ho	(p-) hondje verdwijnen in d	BRUG	(p-) ren		

# Focus on LAT apps: ANNEX

The screenshot displays the ANNEX interface within a Mozilla Firefox browser window. The browser's address bar shows the URL: `http://corpus1.mpi.nl/ds/annex/protected/no_auth.jsp?sessionId=D4553EB501F22C476BC55B46B0263`. The page title is "MPI - ANNEX Interface".

The interface includes a navigation menu with "ANNEX", "manual", and "settings" options, along with a user status of "user: anonymous" and a "login" link. The main content area is divided into several panels:

- Text:** A sidebar menu with options for "Text", "Grid", "Subtitle", and "Timeline".
- Video display:** A video player showing a scene with several people sitting on a wooden platform. It includes a play button, a progress bar, and control buttons (stop, play, next, previous, full screen). A "Settings" button is also present.
- Media information:** A panel displaying metadata:
  - Resource: MISE-1.eaf
  - Media file: MISE-1.mp4
  - Elapsed time: 00:01:42:000
  - Selected chunk: Begin time: -, End time: -, Text: -
- Mini Data Frame:** A panel with a "Tier" dropdown set to "none" and a "Font size" dropdown set to "14".
- Timeline:** A large panel showing a timeline with a red vertical line at 00:01:42:000. The timeline displays subtitles for different tracks (ref1, mb1, gl1, ft1, ref2, etc.) with corresponding text segments. For example, the "ref1" track shows the subtitle: "A nata a musa natabwabogwa kudoki ameya itokulati wala kweni".

At the bottom of the interface, there are control buttons for "Play selection" and "Clear selection", navigation buttons (|<, >|, <<, >>, <, >, +, -), and radio buttons for "Play screen by screen" (selected) and "Play continually". A "Tier text font:" dropdown is set to "Arial Unicode MS".

The status bar at the bottom indicates "Transferring data from corpus1.mpi.nl..."

# Focus on LAT apps: LEXUS

Mozilla Firefox

File Edit View History Bookmarks Tools Help New Window

Back Forward Reload Stop Home

Getting Started Latest Headlines Language Archiving ... LAMUS - Language A... IMDI Browser

Lexus View Help

## Wichita corrected glottal stop

List Tree Selections

StartLetter:  Lexicon:

Selection:

List items

- hahr
- hahr□a
- hahri**
- hahris
- hahwi:ri
- hak
- hak□a
- hak□a:c

Remove New

Result 12 of 2577

First Previous Page 28 Next Last

Lexical Entry View Lexical entry Search

**hahri** vis  
*be angry*

01  
**ne?ati:cháris**  
*he is angry*  
ne?a? ta i2 uc hahri s2  
bad pres pfocpat prevdat angry impf

02  
**ne?ataki:cháris**  
*I am angry*  
ne?a? ta ki? uc hahri s2  
bad pres oneob prevdat angry impf

03  
**ne?ah ke:ki:cháris**  
*I will be angry*  
ne?a? ke? ki? uc hahri s2  
bad fut oneob prevdat angry impf

04  
**ne?ah ke:ki:cha:r?i**



# Focus on LAT apps: VICOS

ViCoS - Visualising Conceptual Spaces

http://lux16.mpi.nl/mpi/vicos/ViCoS.html

ViCoS - Visualising Conceptual ... | http://lux16.mp...ojo/Lexus.html | ViCoS Editor and Navigator | ViCoS Image Manager

ViCoS - Max Planck Institute For Psycholinguistics - http://www.mpi.nl

Lexicon: Yeli\_Dnye | Data Categories: lexeme | Help | RT Mgr

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z te search


< >


Word list view

- chaa pōpu
- chuu kigha
- ch:oo
- chuu
- chile
- cheete
- chii iêê
- chimo
- chalê
- chââm:a
- chu
- chaa pyu
- chêêfi
- chimi chapī
- chala
- chili
- chaa

Welcome | tpile tp:oo | ch:oo | te

Lexical entry view for ch:oo (Yeli\_Dnye)

Lexical Entry	description	description	image	description	image	lexeme	description
[ch:oo]	small rafter	hitching pole for canoes		fish sp. (Orangespine unicornfish, Naso literatus)	null	ch:oo	tree type

Source: 

Relation Type: is\_a\_kind\_of

Target: [tpile tp:oo]

Buttons: Save, Clear

Clear

Add to Knowledge Space | Show in Knowledge Space | Edit in Image Mapper | Edit in Text Mapper

Buttons: [ ] [ ] [ ] [ ]

# New: Portals & Special GIS Stuff

The image displays several web portals and GIS applications. The top left screenshot shows the 'Dane-zaa Community Portal' with a navigation menu:

Stories	Learn about ...	Materials	The Archive
Personal history	Moosehide	Movies	Searching
Traditional stories	Drum	Clickables	Navigating
Animals	Dry meat	Dictionary	Download
Place	Handgames	Phrasebook	Tours
	Snowshoes	Calendar	Google Earth
	Bannock	Alphabet	Studies
		Posters	

The top right screenshot shows a Google Earth interface with a map and an information window titled 'how a got its name, etc'. The bottom left screenshot shows a 'Request Page' in Microsoft Internet Explorer. The bottom right screenshot shows another Google Earth interface with a map and an information window titled 'Ntuan'.

- fostering the creation of special web-sites by REST interfaces and templates
- fostering the GIS presentation by special converters



# New: Cultural Documentation

The screenshot displays the LEXUS Knowledge Spacer interface in Mozilla Firefox. The main window is titled "The LEXUS Knowledge Spacer - Mozilla Firefox" and shows a workspace with several tabs: 'eka kira', 'enana he'e', 'arahu', and 'aati'. The "List of noun words" section on the left contains a grid of words like 'ena', 'ama', 'apau', etc. The "Knowledge Space Construction Area" features a central hub-and-spoke diagram with 'COCONUT' at the center, branching into categories like FOOD, ARTWORK, and MEDICINE. The "Multimedia View Area" at the bottom left shows a coconut image. On the right, a "Lexical Entry" for 'ti' is displayed, including its definition (shell, cowry) and a photograph of a cowry shell. Below this, a video player shows a person using a tool to process a coconut. A timeline at the bottom of the video player shows a duration of 00:01:48:22. In the bottom right corner, there is a small inset image of a group of people gathered around a table, engaged in a collaborative activity.

- collaborative knowledge space with culturally relevant concepts for semantic navigation where concepts are center points for all sorts of information
- genealogy view to come next



## Future evolutions: CLARIN

- Building the next-generation eHumanities infrastructure:
  - (No new data,) but make the existing data interoperable
- Some pillars:
  - Component-based metadata
  - Data Category Registries
  - Persistent Identifiers + Virtual Collections
  - Federated Identity management
  - Simplified license templates



## Future evolutions: AVATECH

- Advancing Video/Audio Technology in Humanities Research
- MPI Halle + MPI Nijmegen + Fraunhofer institute
- Integrate audio/video recognizers in LAT framework

Questions?

# Backup Slides

# Archive Content: Yéî Dnye (Rossel Island)

## Multimedia Lexicon

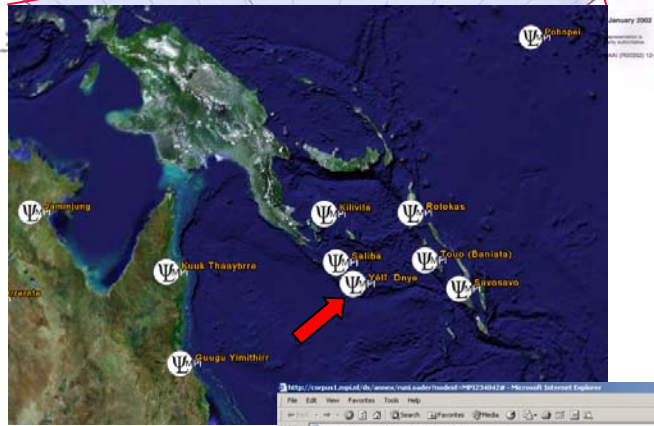
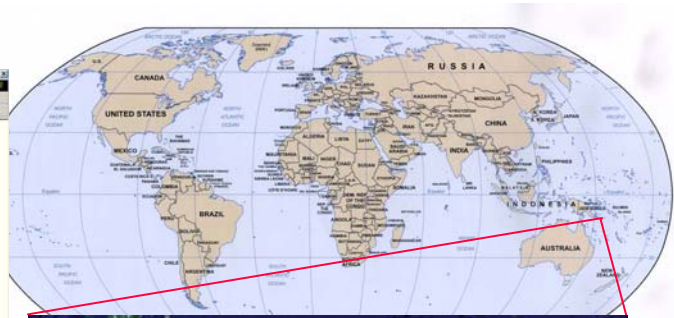
**Browsable Corpus**  
housed by the MPI for Psycholinguistics  
(HTML version)

The Browsable Corpus gives you access to the IMC2-corpora housed at the MPI for Psycholinguistics. It allows you to browse through their hierarchical tree structure, whereby each node in the structure gives you access to further nodes. The nodes are characterized through metadata information, and they often have additional files associated with them: information files and, at the lowest level in the hierarchy, audio/video and annotation files. Metadata and information files are openly accessible, while audio/video and annotation files are usually protected for ethical and legal reasons. If you want to access protected files, please read the accompanying metadata files: they disclose the necessary contact information.

The Browsable Corpus is a preliminary test version for accessing XML-based IMC2-files with normal web-browsers. If you notice any errors, please contact the [Corpus@mpi.de](mailto:Corpus@mpi.de) City browsers newer than Netscape 4.7 are supported.

If you don't see a tree on the left side of the page you may try our [alternative](#) of the IMC2 browser. For displaying the tree on the left side you need to have Java (J2SE) installed ([link](#)).

### Described Corpus



**(WORKSPACE) LEXUS-Lexicon Scheme Viewer**

**Sample lexicon for Rossel Demo 10/Feb/2006**

Lexicon: tpile tree

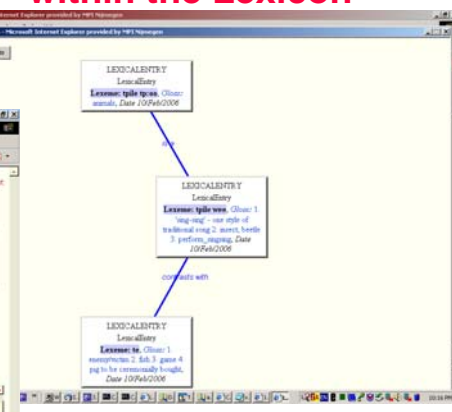
Class: 1 'ing-ang - use style of traditional song, 2 assert, berde, 3 perform\_singing

Example: 1 there's no singing, 2 tpile we las toklokw

**Free Translation:**

Note: 1. M: clear relationship to berde! The song style contrasts with yaa, st and all other styles both musically and lexically. Tpide were are only sang by men and boys, in performances that last all night. Performers wear special grass skirts and carry spears. 2. includes cockroach, mag berdes, wasps, ants, mosquitoes, white ants (and anything with ... we) but excludes bees, bees, case-wants.

### Typed Relations within the Lexicon



### Video Clips



### Photos

**Workspace**

File: r03\_v20\_s5.nsf (MP12340424#)

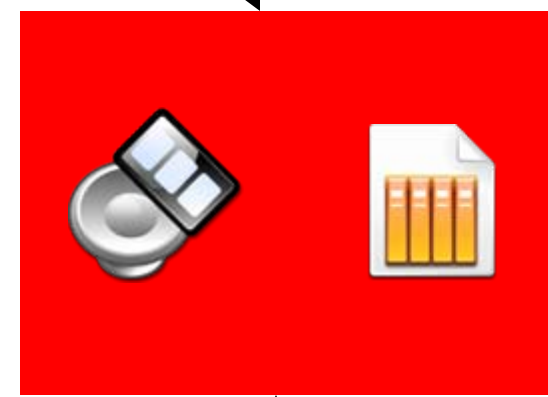
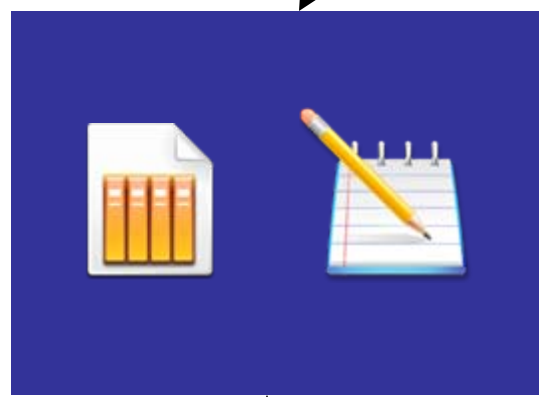
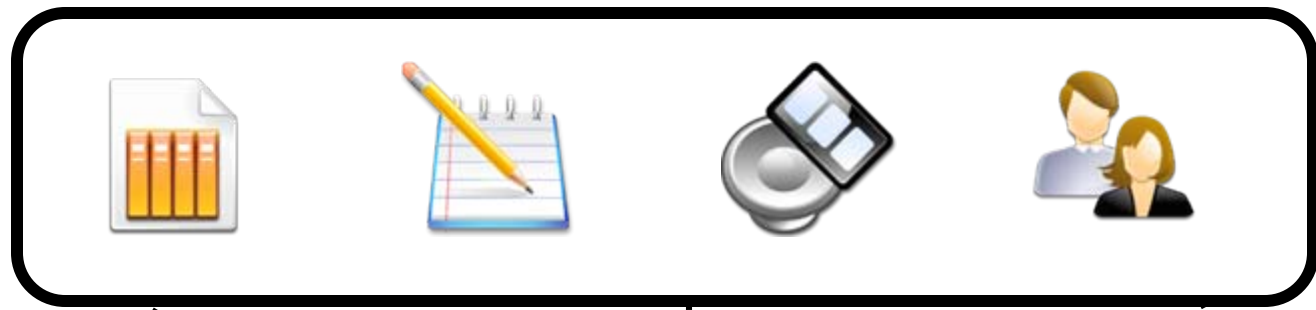
Class: 1  
File times:

00:00:00 - 00:00:00 00:00:00 00:00:00

just sing one! I also have never heard it (tricking?) **was singing** think it will go down here (in recorder) that's where you started at the beginning of the mbwaa. OK, you started to sing Vi's boat's opening it at K. I also made a mistake: because of such a crowd at down, respect seems to me just the same (as Myk's tpile we) at down please I know the man I made a mistake, I went across. Because there were a lot of when you started the beginning of the mbwaa the second one. How people overtake me I never finished it we wanted to see you, you sing many have you in memory? this one is second? You sing Myk's mbwaa that one sing the beginning part, start at the beginning I won't sing it. first. This thing isn't just anything, it's going to record it's going to record people are coming closer. No, just sing it differently (not loud) you go them. This the time for singing the mbwaa, it's coming out of mouth. I away, you Yabénié person. OK, I haven't heard it, sing it in my hearing might sing it in church. You are "drinking/singing the Tpyipi pyu tpile That the beginning of the Vi's? That's the beginning of the mbwaa section (indrogiti) I started doing the Myk's tpile we again because why? I might sing it now you say it properly, he might record it. He's still here I am saying Fiumong as son of Kpas at Wulé. yes, I was eating going down (to foreign parts) do you drink Vi's mbwaa first if I started two pieces and sat water to complementary foods no other boats have on the tpile we. When Kipé's mbwaa first if I started. Tembil. He went right around

### Annotated Media

# Component-based metadata

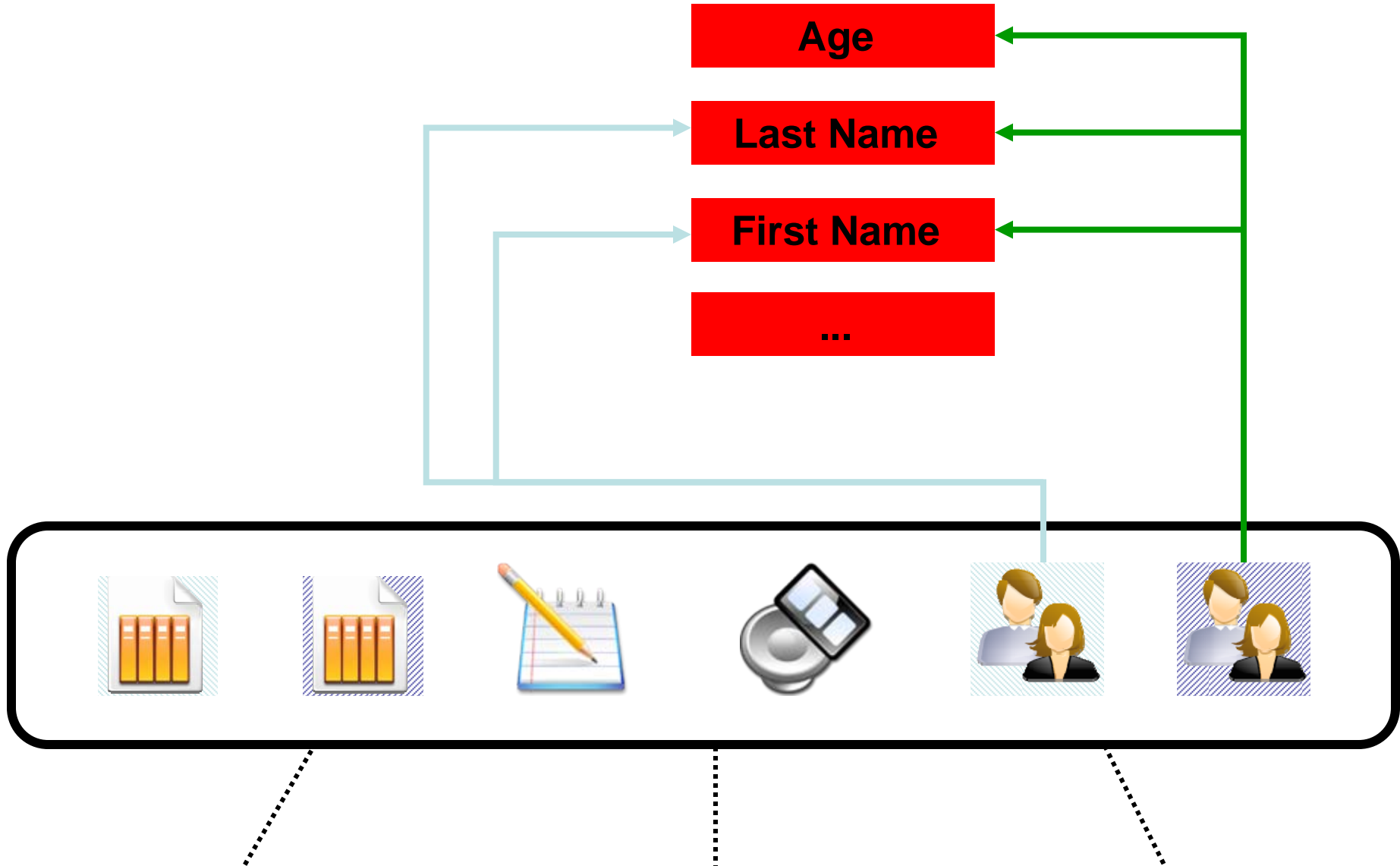


project 1

project 2

project 3

# Data Categories







## References

- LAT: <http://www.lat-mpi.eu>
- MPI Nijmegen: <http://www.mpi.nl>
- CLARIN: <http://www.clarin.eu>
- CGN: <http://lands.let.kun.nl/cgn/>

# Language Archiving Technology

<http://www.lat-mpi.eu>