Online Resources on Oceanic Languages

Jacqueline Ringersma
Max Planck Institute for Psycholinguistics
Online Resources on Oceanic Languages

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?

LAT – Language archiving technology
www.mpi-lat.eu
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
Annotated Media

Video clips, sound files & images

Described Corpus

Multimedia Lexicon

Typed Relations within the Lexicon

Teaser
1. 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
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

http://corpus1.mpi.nl
Clear principles: Data organization and access infrastructure

Browsing the data: IMDI browser

Tree browsing
Clear principles: Data organization and access infrastructure

Browsing the data: IMDI browser

Metadata
Clear principles: Data organization and access infrastructure

Browsing the data: Metadata search
Clear principles: Data organization and access infrastructure

Browsing the data: Metadata search
Clear principles: Data organization and access infrastructure

Browsing the data: resources
AMS on resources in IMDI browser

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

Default setting after upload
Viewing the data: ANNEX viewer
Different ways of accessing:

Community portals

Google Earth Layers
2. Data on Oceanic languages in the MPI archive and how to access
Papua New Guinea:

1. Saliba language data (text, sound, video, annotations)
Papua New Guinea:

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

**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 resources` option
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 resources’ option.
- Some are closed – Use request access option.
3. Other online resources
LEXUS & ViCoS: Web based lexicon tool, multimedia encyclopedia
Other online resources

LEXUS & ViCoS:
Marquesan
Tuamotuan
Yeli Dnye
Teop
Saliba
Online Resources on Oceanic Languages

4. Why archiving?
Why archiving?

Misconceptions about archiving

1. Your stuff is buried here and gone forever

(from Andrew Garret Berkeley Archives)
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)
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"
(attributed to Rufus Pollack))
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 lose 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:
jacquelijn.ringersma@mpi.nl
paul.trilsbeek@mpi.nl