Creating alternative access layers to the DOBES archive from existing metadata structure

Paul Trilsbeek
*Max Planck Institute for Psycholinguistics*

Gabriele Müller
*University of Münster*

Julia Colleen Miller
*University of Washington*
Structure of the Talk

• Intro DoBeS Archive Structure

• Search/Browse Options

• Potential Problems for users

• Beaver Community Portal

• The Technology

• Further Entry Points into the Archive
• Endangered Languages documentation program funded by the German VW foundation

• Started in 1999 with a few pilot projects, until now 50 large documentation projects have been granted

• Central archive for all projects at Max Planck Institute in Nijmegen, about 45 Terabytes of data
• browsing by clicking on the nodes
• metadata search
• annotation content search
DoBeS Archive

ISLE Metadata Initiative

Session
Name: chickadee002
Title: story about chickadee
Date: 2006-02-15

Description
story about chickadee

Location

Project: Beaver

Content

Actors
Actor: Gabriele Mueller
Actor: 891
Actor: 908

MediaFile

WrittenResource

Source

Anonyms

References
Metadata search

13 matches within 670 selected sessions found.

- IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/elicted/word lists/animals/moose_bodyparts001
- IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/elicted/word lists/bodyparts003
- IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/natural use/monologue/narratives/biographic & historical/mooses_hunt001
- IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/natural use/monologue/narratives/place/share_hill_ka-001
- IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/natural use/monologue/narratives/place/share_hill_scenery-001
- IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/natural use/monologue/narratives/creatures/share_hill_fa-001
- IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/natural use/monologue/narratives/traditional stories/mooses_story001
- IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/natural use/monologue/procedural/berry_picking001
- IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/natural use/monologue/procedural/out_mead001
- IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/natural use/monologue/procedural/out_meat002
- IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/natural use/monologue/procedural/make_fleish001
- IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/natural use/monologue/procedural/moose_hide002
- IMDI-corpora/Endangered Languages/DoBeS archive/Beaver Archive/Data/linguistic/natural use/monologue/procedural/moose_hide003
Our stuff is sitting back there" (said the wolf) "go and get it!" and then what?",
DoBeS Archive: Potential Problems

- only in English
- very intricate metadata structure is only useful for specialist users
- not possible to know what is there at a glance
- too many clicks to find something
- community members will be easily discouraged to explore corpus

⇒ the data in the archive should be available to community

How to make access as easy as possible?
The Idea: Community Portals

- provide a webinterface
- keep interface simple and attractive
- with focus on culturally important topics
- choose data in the archive that is relevant
- sort thematically
- use local language
- keep maintenance low, once the portal is set up
- new relevant resources should automatically appear
Dane-zaa Community Portal

This page is created for the members of the Dane-zaa community to facilitate the use of the archive collected by the DoGoS team together with the elders.
Dane-zaa Community Portal

This page is created for the members of the Dane-zaa community to facilitate the use of the archive collected by the DoBeS team together with the elders.

Dane-zaa Community Portal

Learn about...

Drum

Listen to drumming, learn about the different parts of a drum.

paradigm_drum001 - paradigm drum

paradigm_drum001.m4a

paradigm_drum001.wav

paradigm_drum001.txt

paradigm_drum001.pdf
Welcome to OLCAP.

Click on the image to the right to go to the portals.

Click here to go to the main AIATSIS website.

Map copyright Geoscience Australia 2008.
• Technology behind the portals needs to be generic enough to allow for a large variety of portals to be built with the same framework

• Design, layout and selection of search criteria need to be provided by the researchers in collaboration with the speech community

• The archiving team in Nijmegen technically implements the portal and hosts it (if desired)
• REST (REpresentational State Transfer) Web Service
• Query the metadata database via a URL containing a number of parameters
• Search result comes back as XML, which can be further processed inside a dynamic website
• New data that is added to the repository and matches the search criteria automatically shows up in the portal
Example:

http://corpus1.mpi.nl/ds/imdi_search/servlet?request=getMatches&num=50&query=Session.Name="*drum*"&type=advanced&nodeid=MPI79025%23&includeTitle=true&includeUrl=true&includeResources=true&returnType=xml
Dane-zaa Community Portal

This page is created for the members of the Dane-zaa community to facilitate the use of the archive collected by the DoBeS team together with the elders.

Drum

Listen to drumming, learn about the different parts of a drum.

- paradigm_drum001 - paradigm drum
  - paradigm_drum001.m4a
  - paradigm_drum001.wav
  - paradigm_drum001.txt
  - paradigm_drum001.pdf
Plone CMS

- Standard, open source Web Content Management System
- Written in Python
- Large number of add-on modules
- Built-in support for XML processing
- Written a custom content type for Plone: IMDI-query. Sends a query to the LAT server, result comes back as XML and is parsed into a nicely formatted list of resources
Tier 1: टेलीटेल, बाल्किपोलदे बरिता आउँदा दाख्चे?

Tier 2: Coming from the Waling settlement to this side to the Tele settlement.

Tier 3: [No source tier selected]

Tier 4: [No source tier selected]
### Timeline

<table>
<thead>
<tr>
<th>Language</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>cmp@SR</td>
<td>You were living over there, you were living at mulgaun?</td>
</tr>
<tr>
<td>nep@SR</td>
<td>अपनी गांव में रहा था, उसके नज़ारे में रहा था।</td>
</tr>
<tr>
<td>gran@SR</td>
<td></td>
</tr>
<tr>
<td>ext@SR</td>
<td></td>
</tr>
</tbody>
</table>
...nochmol waut vetahle von, um, waut ekj aus Kjind erläwt hab.
...tell something again about, um, what I experienced as a child.

Daut wea, uh, em Somma emma sea woam.
It was, uh, always very warm in the summertime.

Un, uh, uh, dee,
And, uh, uh, the,

wi haude väl Oft.
we had lots of fruit.

Miene Taunte haud sea väl Oft jëdreegt.
My aunt had dried a lot of fruit.

Dee deed emma väl Oft dreeje,
She always dried a lot of fruit,
Google Earth

Snare Hill

Madânts’atl’oje “Snares Set Around It”

Sammy Alslo shares a story about the importance of Madânts’atl’oje to the Dane-zaa people. The story is told in the Dane-zaa language, with English subtitles.

Directions: To here - From here
Conclusions

• Different user groups have different requirements regarding interfaces

• The same archival material can be presented in various ways, especially when making use of XML based formats

• Web Services make it easy to use archived material in different online contexts in a programmatic way