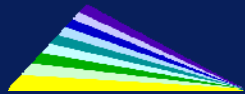


Arbil



Jacqueline Ringersma



Arbil

Arbil aim:

- Design a corpus lay-out

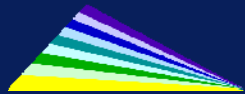
- Add IMDI metadata files and resources to the corpus

- Modify and enrich existing IMDI metadata files

Arbil is designed so that it can be used offline in remote locations.

- Create a local corpus according to IMDI standards

When an internet connection is available existing corpus branches can be downloaded into Arbil or locally created corpora can be uploaded to the archive



Arbil

Arbil is also the archive GATE KEEPER

Check if a file is an archivable format

Will reject resources which are 'unarchivable' to be uploaded

Archivable:

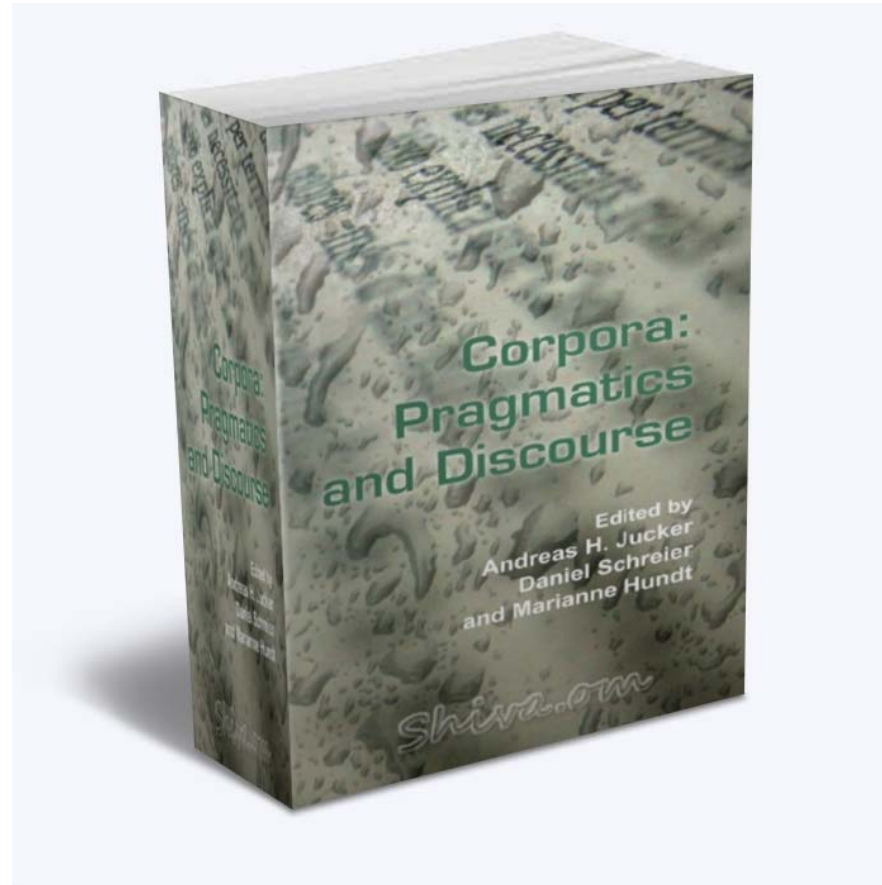
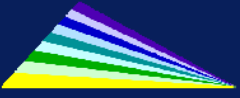
Audio: wav, m4a

Video: mpg, mp4

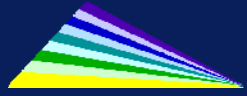
Text: ELAN, toolbox, XML (LEXUS)

Non-archivable:

Msoffice files (or other proprietary formats)



What is a corpus?



What is a corpus?

A collection of samples of written and/or spoken language

Structured in an appropriate manner

Resources described with metadata (session resource bundles)

What is a corpus?

IMDI Browser - Mozilla Firefox

File Edit View History Bookmarks Tools Help New Window New Tab AiOS-Toolbar IE View

Back Forward Reload Stop Home http://corpus1.mpi.nl/ds/imdi_browser/ ☆ Google

Firefox Links VPN-MPI FF Start surfsafe Welcome to the Max P... Gmail - Inbox - jringer... MPI mail

IMDI Browser

IMDI-Browser [about](#) user: anonymous [login](#) [logout](#)

IMDI Metadata Domain

The IMDI Metadata Domain allows you to browse and search in the whole domain of linked IMDI metadata descriptions as they are registered at the IMDI portal at the MPI for Psycholinguistics. All Metadata descriptions are openly accessible, for many resources however one needs to ask access permission. Read further to find out how this is done.

Please note that many corpora are hosted at other resource centers. If a remote server is down, those resources cannot be accessed even though they are still listed within this portal.

How to explore the archive:

On the left hand side, you see the whole hierarchy of linked IMDI corpora, which you can explore by clicking on the small circles or by double-clicking the names. If you click on an item with the **right mouse button** (or control-click with a single button Mac mouse), you will get a contextual menu with additional functionality, depending on what kind of item you have selected:

- view node
- create bookmark
- metadata search
- annotation content search
- set access rights
- download
- request resource access

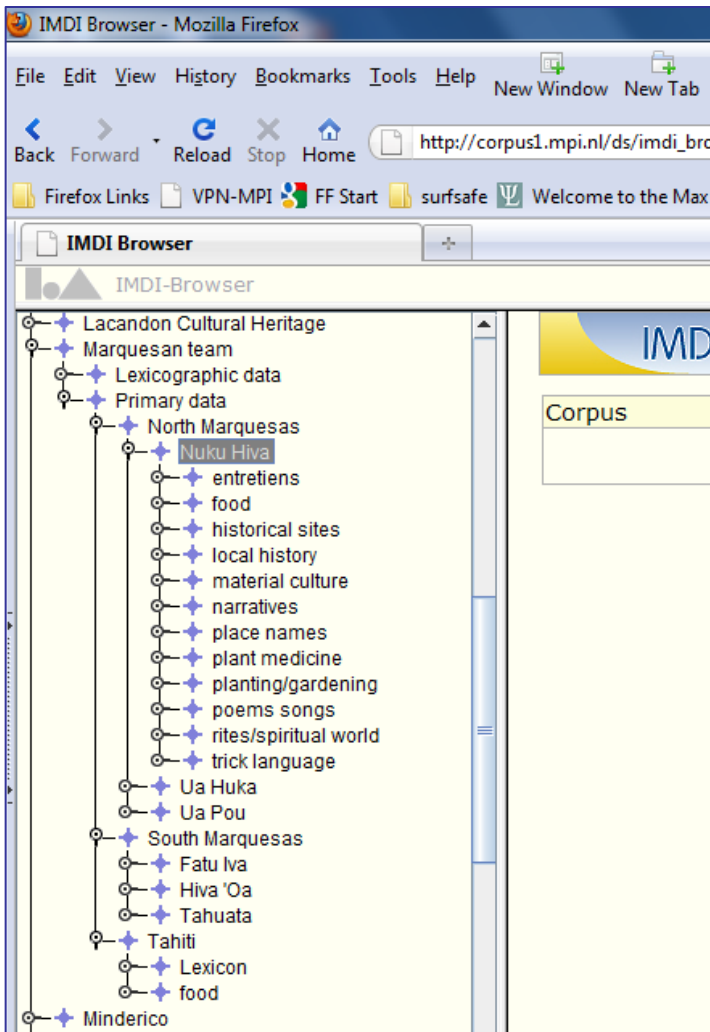
• **view node:** shows the content of a file. For Metadata files, you will get the content of the metadata description. For media files,

Done

Arbil-installation-... IMDI Browser - M... Microsoft PowerP... manual-arbil.pdf ...

EN 2:53 PM 2/02/2010

What is a corpus?



IMDI Browser - Mozilla Firefox

File Edit View History Bookmarks Tools Help New Window New Tab

Back Forward Reload Stop Home http://corpus1.mpi.nl/ds/imdi_bro

Firefox Links VPN-MPI FF Start surfsafe Welcome to the Max

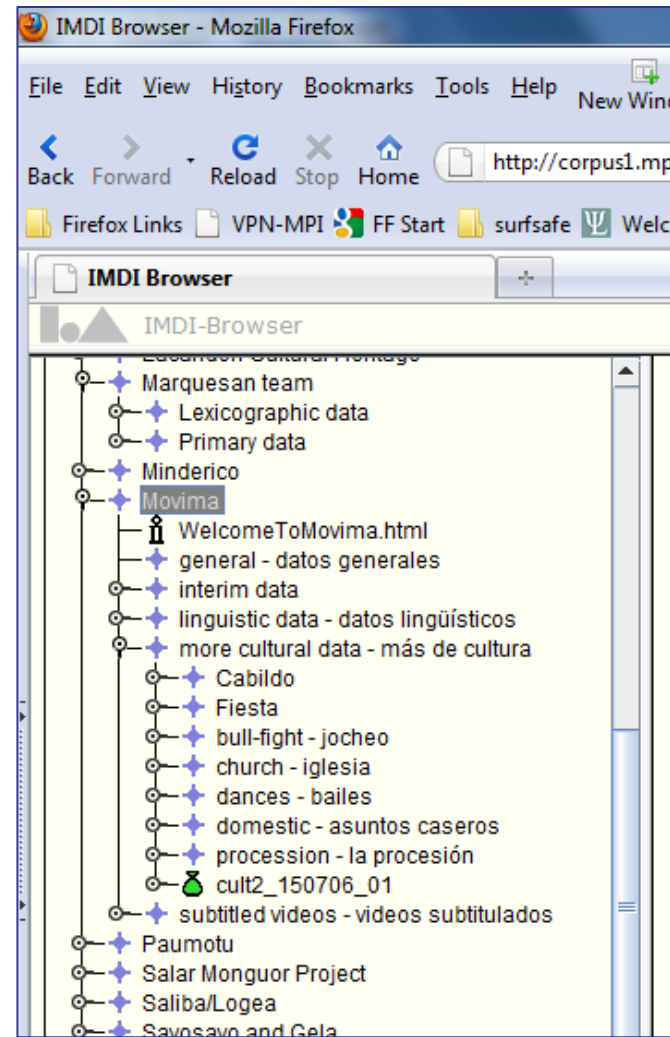
IMDI Browser

IMDI-Browser

- 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
 - Minderico

IMD

Corpus



IMDI Browser - Mozilla Firefox

File Edit View History Bookmarks Tools Help New Window New Tab

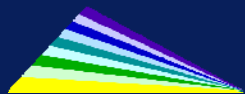
Back Forward Reload Stop Home http://corpus1.mpi.nl/ds/imdi_bro

Firefox Links VPN-MPI FF Start surfsafe Welcome to the Max

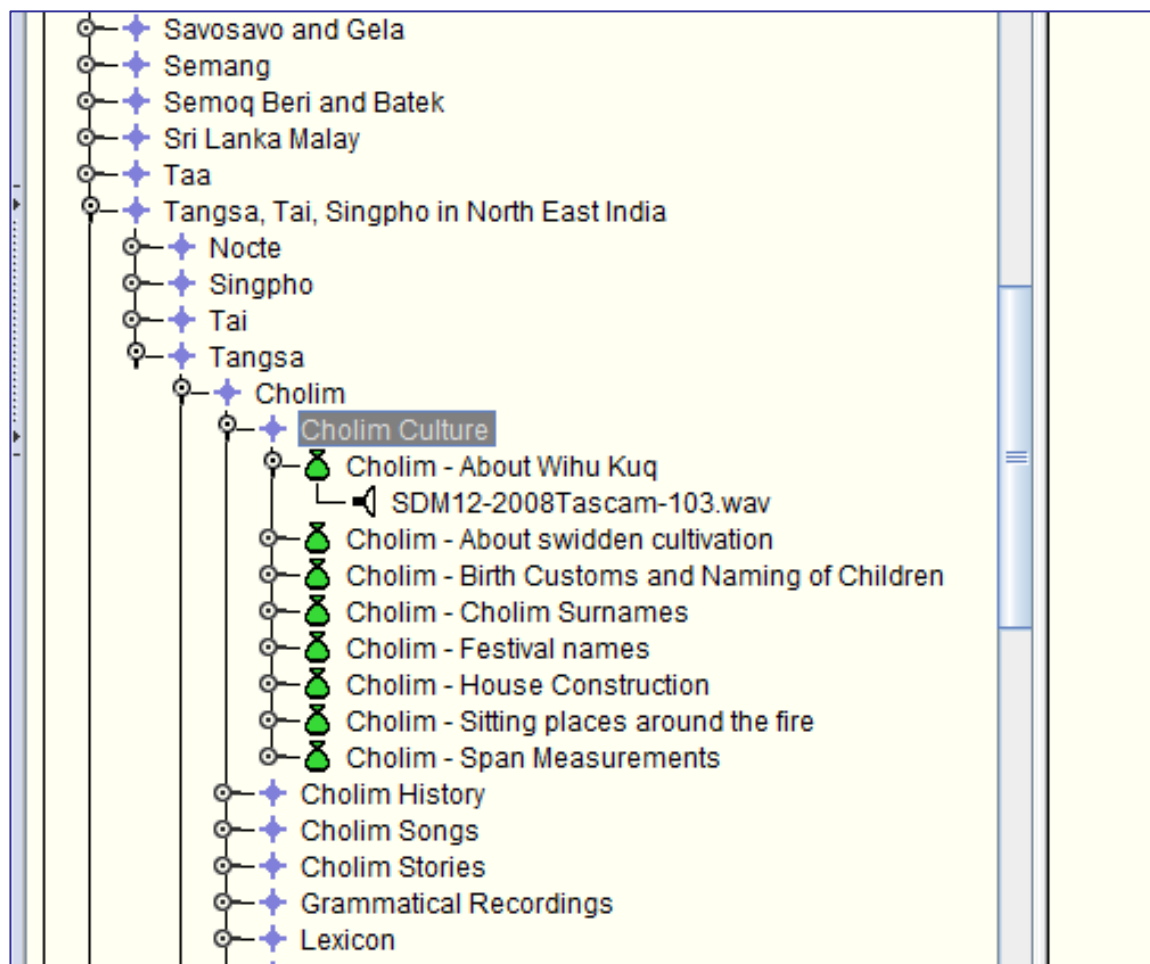
IMDI Browser

IMDI-Browser

- Lacandon Cultural Heritage
- Marquesan team
- Lexicographic data
- Primary data
- Minderico
 - Movima**
 - WelcomeToMovima.html
 - general - datos generales
 - interim data
 - linguistic data - datos lingüísticos
 - more cultural data - más de cultura
 - Cabildo
 - Fiesta
 - bull-fight - jocheo
 - church - iglesia
 - dances - bailes
 - domestic - asuntos caseros
 - procession - la procesión
 - cult2_150706_01
 - subtitled videos - videos subtitolados
- Paumotu
- Salar Monguor Project
- Saliba/Logea
- Savosavo and Gela



What is a corpus?



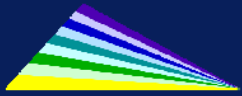


The screenshot shows a Mozilla Firefox browser window with the following details:

- Address Bar:** <http://www.lat-mpi.eu/tools/arbil>
- Page Title:** Arbil — Language Archiving Technology
- Navigation:** Home, Tools, Addit, AMS, Annex/Trova, **Arbil**, Arbil-forum, Imdi Browser, Elan, Imdi, Lamus, Lexus, Synpathy, Vicos, Papers, Status, Related Links.
- Page Content:**
 - Arbil**
IMDI Metadata Editor, Browser & Organizer Tool
 - Recommended Version (stable)**
Run the recommended version via webstart
 - Testing Version (has newer features but is not fully tested)**
Run the testing version via webstart | Download the debian linux package for the testing version | Download the Windows installer for the testing version
 - Manual**
Full manual (Online) | Full manual (PDF) | Presentation (PDF) | A4 Guide
 - Support**
Go to the forum
- Right Sidebar (news):**
 - New LEXUS version (2009-12-10)
 - Archive server down-time, Friday December 4 (2009-12-01)
 - New A4 guide image metadata (2009-11-17)
 - new Annex version (2009-11-12)

The Windows taskbar at the bottom shows the system tray with the date and time: 11:20 AM 5/02/2010. Opened applications include MetadataArbil, Arbil, and Microsoft PowerPoint.

www.lat-mpi.eu/tools/arbil



Work Flow

