



# GE: A flexible presentation platform for LR



Alex Dukers

Jacqueline Ringersma



## **Presenting Linguistic Resources**

### **Geographical presentation**

### **GIS platform: Google Earth (why GE?)**

### **The MPI Google Earth overlay**

(how to use it, how to create it, “future” options)

## **Demo**



# Presenting Linguistic Resources

## Search metadata fields:

PARADISEC:

PARADISEC - quick catalogue search

Persistent Identifier

Book name

Language

Country

Length

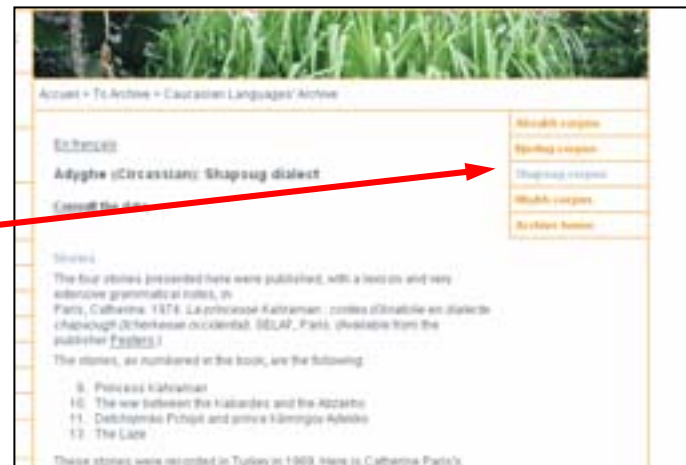
Latitude

You need knowledge on the content!

## Search by links on website:

LACITO:

Ok: for small collections





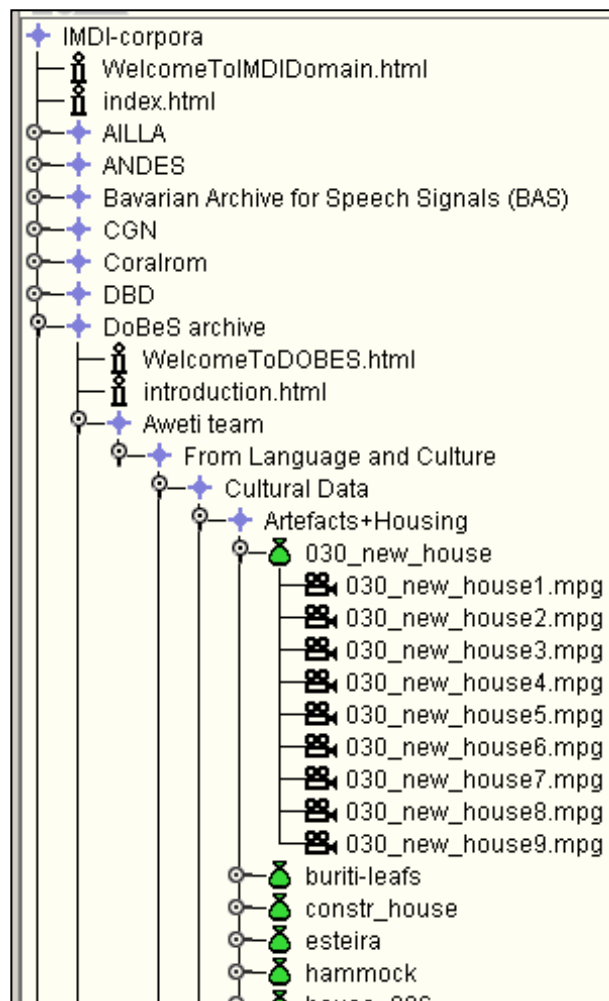
# Presenting Linguistic Resources

## Search metadata plus visual representation:

MPI/DoBeS archive

Still: some knowledge required,  
but user interface is more intuitive

All these presentation methods:  
aimed at content, professional users!





# Geographic presentation

## Geographical search:

The screenshot displays a web application interface for geographic search, divided into three main sections:

- Left Panel:** A sidebar titled "The Archive of the" with a search bar and a list of categories including "Alaska Native", "Africa", "Asia", "Europe", "North America", "South America", and "Australia".
- Middle Panel:** A map titled "Alaska Native" with the instruction "Click on a language name". Below the map, there is a list of language names: "Aleut", "Athabaskan", "Eskimo-Aleut", "Tlingit", "Creek", "Algonquian", "Koyukon", and "Tahltan".
- Right Panel:** A section titled "Browse the Archive" with a navigation bar containing "By: Country | Family | Name | Data Type". Below this, it says "Or explore by:" followed by a world map titled "Geographic Viewer". The world map is color-coded by continent: North America (red), South America (yellow), Africa (orange), Europe (light blue), Asia (green), Australia (yellow), and Antarctica (dark green).

OK, but for now this is limited to individual organisations, each one using its own application.



# Geographic presentation

## **Benefits:**

Overlay over the content

Aimed at general public, convenient and intuitive

PR – can attract new depositors/users

Integrate different research disciplines and draw new correlations between them (e.g. ethnographic & linguistic)

Integration of different language archives in a unified manner







# LR collection

**Intera Slovene - English**

**el da** Evaluations and Language resources Distribution Agency

**Intera**

INTERA (Integrated European language data Repository Area) was a European Project with essentially two pillars :

- (a) to build an integrated European language resource area by connecting international, national and regional data centres
- (b) to produce new multilingual language resources.

Bilingual parallel corpora : The texts belong to the domains of education, health, law, tourism and environment. The parallel texts have been aligned as well as annotated at the structural (sentence level) and linguistic levels (PoS tagging and lemmatisation). For Slovene - English the corpus contains 2 million words per language.

The archive contains a Bulgarian lexicon with 2,052 terms. They are distributed into several domains : Law, Law-Politics, Politics, Education, Environment, Health, Tourism, Finance.

[ELDA Intera project page](#)  
Data was collected by LSP ( Institute for Language and Speech Processing) [LSP website](#)

[MPI Intera project page](#)  
[Browseable corpus Slovene - English](#)

Directions: [To here](#) - [From here](#)

©2006 Europa Technologies  
Image © 2006 NASA  
Image © 2006 TerraMetrics  
Image © 2006 GeoContent

Streaming 100% Eye alt 1186.03 mi





# LR collection

IMDI-Browser about user: anonymous logo

Slovene - English

**IMDI** ISLE Metadata Initiative

**Session**

**Name** sve2  
**Title**  
**Date** 2004

[Location](#)

[Project](#) **INTERA**

**Name** INTERA  
**Title** INTEgrated European language data Repository Area  
**Id** 22076Y2C2DMAL2

[Contact](#)

**Description**

INTERA was a European Project with essentially two pillars : (a) to build an integrated European language resource area by connecting international, national and regional data centres, and (b) to produce new multilingual language resources. The first goal involves the integration of a critical mass of different types of language resources with the help of metadata descriptions and the interlinking of the resulting distributed resource repository with an existing tool repository thus enabling users to directly start suitable tools on the included resources. This integrated and interlinked metadata description domain has been aimed at facilitating the access to language resources in Europe, especially for professionals in industry, the eContent business, research and education, thus increasing the usage of the resources

**Keys**

**Intera.Domain** Law

[Content](#)

**Actors**

[WrittenResource](#)

**ResourceLink** svez-sl.txt  
**MediaResourceLink**  
**Date** Unspecified  
**Type** Primary Text  
**SubType** Unspecified  
**Format** text/plain  
**Size** 2.000.000 words

[Validation](#)

**Derivation** Unspecified  
**CharacterEncoding** ANSI  
**ContentEncoding**  
**LanguageId** RFC1766:x-sll-SLV  
**Anonymized** Unspecified

[Access](#)



# LR collection

The screenshot displays the Google Earth interface with a map of Indonesia. The left sidebar is open to the 'Places' panel, showing search results for 'Yeli Dnye' and other languages. The map shows several locations marked with icons and labels: Lamu-lamu ('umpuykamu dialect'), Kilivila, Saliba, Yeli Dnye, Yele, Teop, Rotokas, Touno, and Savosavo. The bottom status bar shows coordinates: Point: 8127.53 837.5, 151.25 21.49° E elev: 106 ft. The Google logo and copyright information are visible in the bottom right corner.

Places

- [Yeli Dnye](#)
- [Turkish](#)  
Gesture project by [A. Ozgenel](#) on Turkish children and adults. The corpus contains
- [Tidore](#)  
Language of Tidore, Halmahera Tengah, Indonesia. Tidore is classified as an
- [Waimai Atrorai](#)  
The Waimai Atrorai language, which they call kinja iara, "people's language," belongs
- [Yelk Dnye](#)
- [Yucatec](#)  
Yucatec (Yucatec Maya) is a Mayan language spoken in the Yucatán Peninsula,
- [Language Acquisition](#)
- [Bonafed corpora](#)
- [Sign language Typology](#)

Layers

- [Layers](#)
- [terrain](#)
- [2006 US Election Guide](#)
- [Featured Content](#)
- [roads](#)
- [borders](#)
  - [International Borders](#)
  - [Country Names](#)
  - [Island Names](#)
  - [Coastlines](#)
  - [Est Level Admin Borders \(States Provi](#)

Image © 2006 NASA  
© 2006 Europa Technologies  
Image © 2006 TerraMetrics

©2005 Google

Point: 8127.53 837.5, 151.25 21.49° E elev: 106 ft Streaming: 100% Eye alt: 1068.83 m

## Yéli Dnye

Pioneers of Island Melanesia -  
a joint project between British, Dutch, German and Swedish teams

Yéli Dnye (also known as Rossel, Yela, Yele, Yelejong, Yeletnye) is a Papuan language spoken on Rossel Island, Louisiade Archipelago, Papua New Guinea. Although surrounded by Austronesian languages, Yéli Dnye shows little evidence of influence by them, making this language an isolate. Yéli Dnye is known as the language with the world's most complex phonemic inventory. Project leader is [Stephen Levinson](#).

[browsable corpus link](#)

[LEXUS example](#)

For this demonstration of Lexus use "demo" as username and "demo" as password.

[ANNEX example 1](#) [ANNEX example 2](#) [ANNEX example 3](#)

[Pioneers of Island Melanesia webpage](#)



Village on Rossel Island



LEXUS and ANNEX examples

Directions: [To here](#) - [From here](#)








link to IMDI web browser:

The image shows a screenshot of the IMDI browser interface. On the left is a tree view showing the hierarchy of data. The root is 'Yale', followed by 'Levinson', 'Cultural data', and 'Photos'. Under 'Photos', there is a 'Fieldtrip by date' folder containing years from 1995 to 2003. The year 2003 is expanded to show a folder 'R03July12\_1', which is further expanded to show a list of files including HTML and JPEG files for each day from July 1st to July 10th, 2003. On the right side of the interface, there is a 'Media' section. It displays three media items, each with a thumbnail image and a title: 'R03July12\_1\_01', 'R03July12\_1\_02', and 'R03July12\_1\_03'. Below each thumbnail, the text 'Description: No description available' is displayed.

**Media**


**R03July12\_1\_01**



**Description:**  
No description available

---


**R03July12\_1\_02**



**Description:**  
No description available

---


**R03July12\_1\_03**





# Link to ANNEX examples:

ANNEX Workspace user: anonymous [logout](#)



0:5 00:22:19.675  
00:22:19.676 - 00:22:20.646 00:00:00.970

File: r03\_v19\_s2.eaf  
combined 14 hide controls hide media hide QT CSV info

FsGest Action FsGest

hide times

3rd direction dfto Points NE Gesture held points E "gone" looks up (at K's roof leaks?) points at K's roofleaks points W index to 2nd finger index to 2nd finger knocks Mgaa holds 3 then 4 fingers 2 fingers 3 fingers counting 2 fingers holding 1 finger points down wordless point apuul hand to mouth touches K [looking in his bowl](#) nods turns to M (nodi) eyes down eyes up

E

M to oo móó u nodu

K ee ki pini voo ki pini mbwémi kn nyáá kwono ala mbwé ala mbwémi ala

Code\_person MD MD

I il il yes ee ee

Interact Building is aligned N-S, with S on the right

Feedback

Repair

Place

DemoQ

Demo\_Repair

notes The conversation between Ian and Michael at Abaleti - this part v19\_s2 beginning concerns how Michael Kim Iliene's wife Nk-akada (after bearing 3 children) was stolen back fro



# Link to LEXUS example:

LEXUS Lexical Resource Search View extra demo

Retrieved 16 entries for **t**  
 Number of entries on page 25  
 Page number: 1

**Lexical Entry**

- tpile tp:oo**== animals
- tpile wee**== 1. 'sing-sing' - one style of traditional song 2. perform\_singsing 3. insect, beetle
- tpile\_pē**== snake (lit. thing\_long)
- te**== 1. fish 2. enemy/victim 3. pig to be ceremonially bought 4. game
- tóótpi**== fish sp.
- tpy:aakaa**== butterfly
- too pene**== fish sp.
- tóódpi**== fish sp.
- tiki**== 1. woman's name 2. bird\_type
- tii**== 1. tea (English loan) 2. lineage, ancestors 3. dry/yellow(leaf) 4. sand 5. bird sp (Paradise kingfisher) 6. relatives (not in-laws) 7. line
- tada**== kingfisher (Halcyon chloris?)
- t:a**== red\_fire\_fish Pterois volitans
- tiye**== 1. bird of paradise 2. village name
- tapē**== gekko/lizard type
- tēēpi**== 1. snake type, python 2. tree sp.
- tēpē we**== ant lion

http://lux07.mpl.nl - Lexus - Mozilla Firefox

LEXUS LexicalEntry Edit View demo

**Lexical Entry:** tpile wee

Description: 1. 'sing-sing' - one style of traditional song, 2. perform\_singsing, 3. insect, beetle

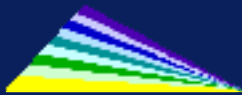
LEXUS

- lexicon
- lexiconInformation
- lexicalEntry
  - ac:tpile we
  - t:
  - note: No clear relationship to beetles! This song
  - note: includes cockroach, stag beetles, wasps, .
  - senset: contrast set (yaa, n̄amé, tpile wee)
  - senset: is-a tpile tooo, contrast set (te 'fish', 'u
  - w: tpile wee
  - d: 10 Feb 2006
  - g: N & V. N takes 2 positionals:
  - e: 'There's no singsing'
  - e: tpile we daa tóó-ko
  - d: 'sing-sing' - one style of traditional song
  - d: insect, beetle
  - d: perform\_singsing
- Media
  - Video<<video>>
  - Audio<<audio>>
  - Image<<image>>

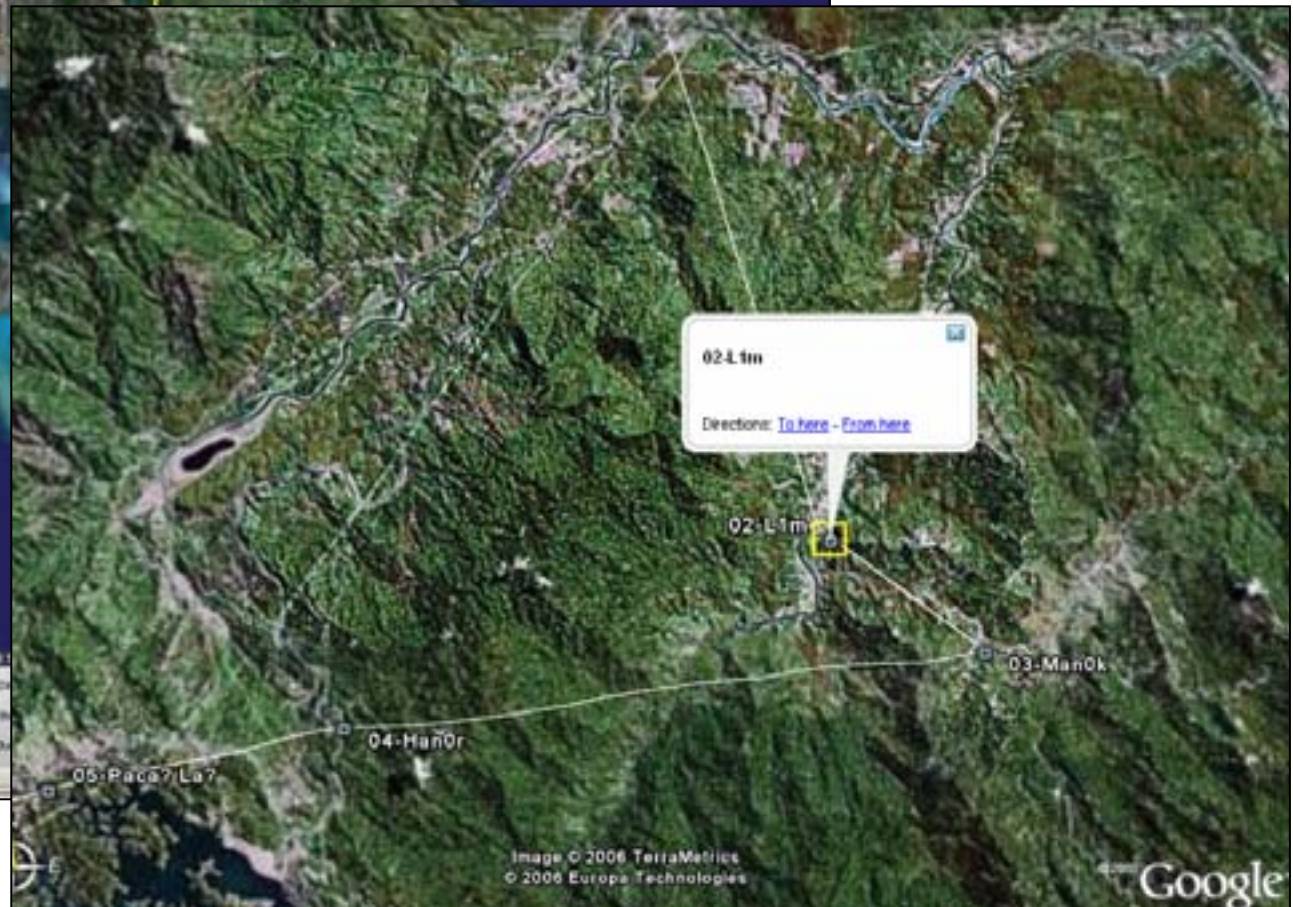
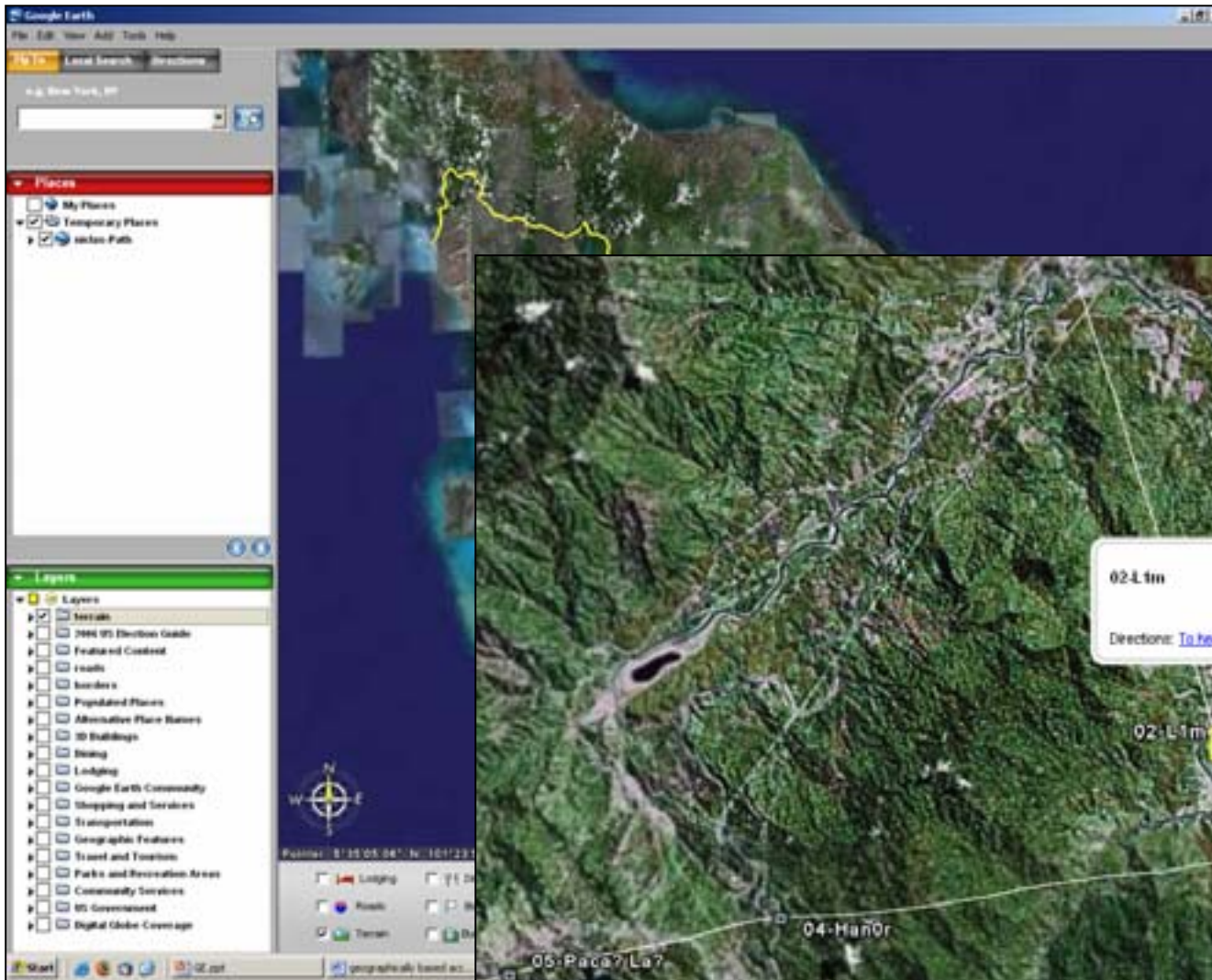
Transferring data from lux07.mpl.nl...





# To be integrated: fieldwork paths





# Why Google Earth

It's XML:



```

<Placemark>
  <name>Venezuela</name>
  <description>
    <![CDATA[<br />
    <br />
    </img>
    <br />
    <br />
    Corpus containing data on the Uruak, Yanomamã ,
    Mapoyo, Makiritari, Panare, Sanumã, Kariana,
    Warao and Yawarana languages of Venezuela.
    Resources include articles, songs, narratives, and wordlists (and other genres).
    <br />
    <a href="http://corpus1.mpi.nl/ds/imdi_browser?rootnode=MPI312876%23">
    browsable corpus link (partially converted to IMDI framework)</a>
    <br />
    <a href="http://www.ailla.utexas.org/search/view_resource.html?country_id=5&name=Venezuela">
    AILLA archive</a>]]>
  </description>
  <LookAt>
    <longitude>-66.64963797841051</longitude>
    <latitude>6.439514949938382</latitude>
    <range>2000000</range>
    <tilt>35.00000000000152</tilt>
    <heading>-6.785927243998965e-014</heading>
  </LookAt>
  <styleUrl>#khStyle1002</styleUrl>
  <Point>
    <coordinates>-66.62484052032704,6.605781049085067,0</coordinates>
  </Point>
</Placemark>

```



# Why Google Earth

- It's XML: other applications, data harvesting
- Freeware, runs under Windows, OS X, Linux
- Large player in GIS presentation platforms
- Fast streaming of data
- Well known to the general public, easy to use
- Lively community, help forum, many plugins
- Regular update of maps, covers the whole globe
- Stand alone application (faster than Java)
- Easy conversion to GoogleMaps



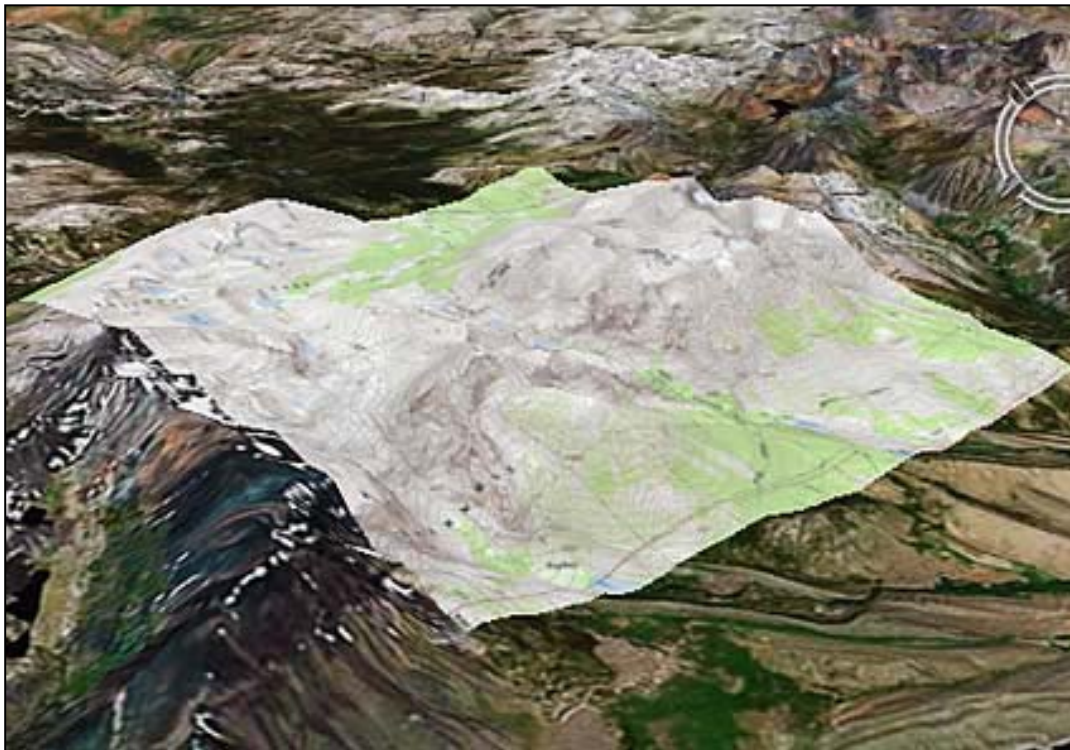


# "Future" options

Technical:

1. Image overlay – static or dynamic

e.g. marking of dialect areas, migration (language) population





# "Future" options

Technical:

1. Image overlay – static or dynamic

e.g. marking of dialect areas, migration (language) population

2. From metadata to Google earth overlay requires coordinates in the metadata descriptions

3. Incorporation of more Language Resources

(you can send us your details! And we will include these in the overlay)



# More information:

[www.mpi.nl/services/mpi-archive/GE\\_language\\_sites](http://www.mpi.nl/services/mpi-archive/GE_language_sites)

Max Planck Institute for Psycholinguistics

## Google Earth Language Sites collection

Up one level

Navigate the 3d globe in Google Earth and get access to various Language Archives.

Use geographic browsing to view a collection of places representing various research locations of the Max Planck Institute for Psycholinguistics and other corpora. New placemarks will be added over time.

Get the Google Earth language sites file.

Download the Google Earth free version here, choose between Windows 2000/XP or OS X. For GE beta 4 there is also a Linux version available. For more information about geographic browsing and how to manipulate the contents of your own placemarks check the primer. Visit the tour on the Google Earth site to learn basic navigation or watch some video tutorials which demonstrate how to add your own data. Please contact [Alex Duker](mailto:Alex.Duker@mpi.nl) to have your placemarks incorporated into the MPI language sites collection or if you have comments on the present ones.

**MPI Links**

- [Description in German](#)
- [Annual Reports](#)
- [Language Sites](#)
- [Browseable Corpora MPI](#)
- [LAN \(Archives Newsletter\)](#)
- [MPI Tools](#)
- [DolBeS](#)
- [Online Experiments](#)
- [F.C. Donders Centre Radboud University](#)
- [More links...](#)

Disclaimer stemap