

User-Experience von Spracharchiven

Eine Neubewertung der Interaktion von Archiv und Nutzern

KA³



GEFÖRDERT VOM
Bundesministerium
für Bildung
und Forschung

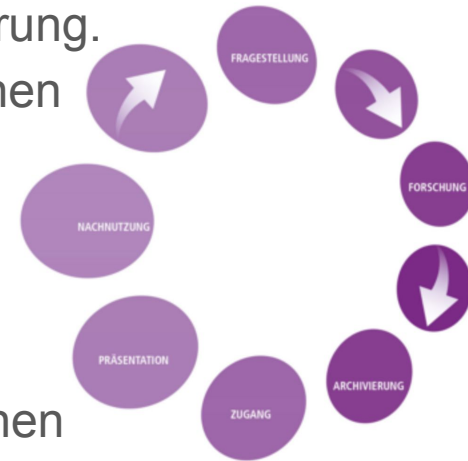
Universität
zu Köln



DCH
Data Center for the Humanities
Köln
Köln
Köln
für die Geisteswissenschaften

Nachhaltige Strategie für Spracharchive

- Spracharchive sind zentral für eine empirische, datengetriebene Sprachwissenschaft.
- Nachnutzung ist essentiell für die Akzeptanz und Finanzierung.
- Spracharchive sind Plattformen, die den Austausch zwischen Forschern ermöglichen sollen.
- Es gibt (mindestens) zwei Nutzerrollen: Produzenten und Konsumenten.
- Produzenten und Konsumenten haben unterschiedliche Interessen, eine zentrale Aufgabe eines Archivs ist, zwischen diesen zu vermitteln.



Warum interessiert uns das?

KA³

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

- Kick-off-Phase von einem großem BMBF-geförderten Projekt:
KA³ - Kölner Zentrum Analyse und Archivierung von AV-Daten
 - Überarbeitung, teilweise Neukonzeption und Verstetigung einer technischen und institutionellen Archivrösung
 - Bereitstellung von Analyse-Tools auf Audio-Daten (z.B. automatische Segmentierung, Sprecherdiarisierung, Pausenerkennung)
- Welche Anforderungen existieren? Von wem werden sie formuliert? Gibt es Widersprüche? Wenn ja, wie kann man zwischen diesen vermitteln?
- Partner:



Data Center for the Humanities
Kölner Datenzentrum
für die Geisteswissenschaften



Max Planck Institute
for Psycholinguistics



Fraunhofer
IAIS



DARE
Digital
Averroes
Research
Environment



Data Center for the Humanities
Kölner Datenzentrum
für die Geisteswissenschaften

Potentielle Aufgaben von Spracharchiven

- Archivierung
- Zugänglich machen
- Kuration
- Publikation von Primärdaten
- Forschungsunterstützung

Nutzung / Nachnutzung

- gute Voraussetzungen, weil die Forschungscommunity Fokus auf die Produktion von Datenkorpora hat
- allerdings: hohe bzw. sehr diverse, oftmals widersprüchliche Anforderungen
- Nachnutzung bleibt hinter den Erwartungen zurück
 - Produzenten: Die Praxis auf bestehende Dokumentationsdaten zurückzugreifen ist noch nicht etabliert genug
 - Produzent/Konsument: die Dienste der Archive sind nicht gut genug
 - Archivar/Konsument: Dokumentationsdaten sind nicht ausreichend aufbereitet, um sie sinnvoll nachzunutzen (Qualität Metadaten, Annotationen)
 - Archivar: die Zugänglichkeit wird durch die Produzenten eingeschränkt

Nutzerrollen

- Daten-Produzent
- Daten-Konsument
- (Öffentlichkeit, Speech-Community, Archivare, ...)

Daten-Produzent

Wissensstand & Interessen

Bedürfnisse-

- Projekt-spezifische Präsentation
- Zugangskontrolle
- Anerkennung des Deposits als wissenschaftliche Leistung

Weitere Differenzierung möglich:

Persona Prof, Persona PostDoc, Persona Doktorand

Daten-Konsument

Wissensstand & Interessen

Bedürfnisse

- Projekt-agnostisch und -übergreifend Datensets finden und zusammenstellen
- Freier Zugang bzw. möglichst geringe Hürden
- will unkompliziert überprüfen können, ob Daten relevantes enthalten könnten

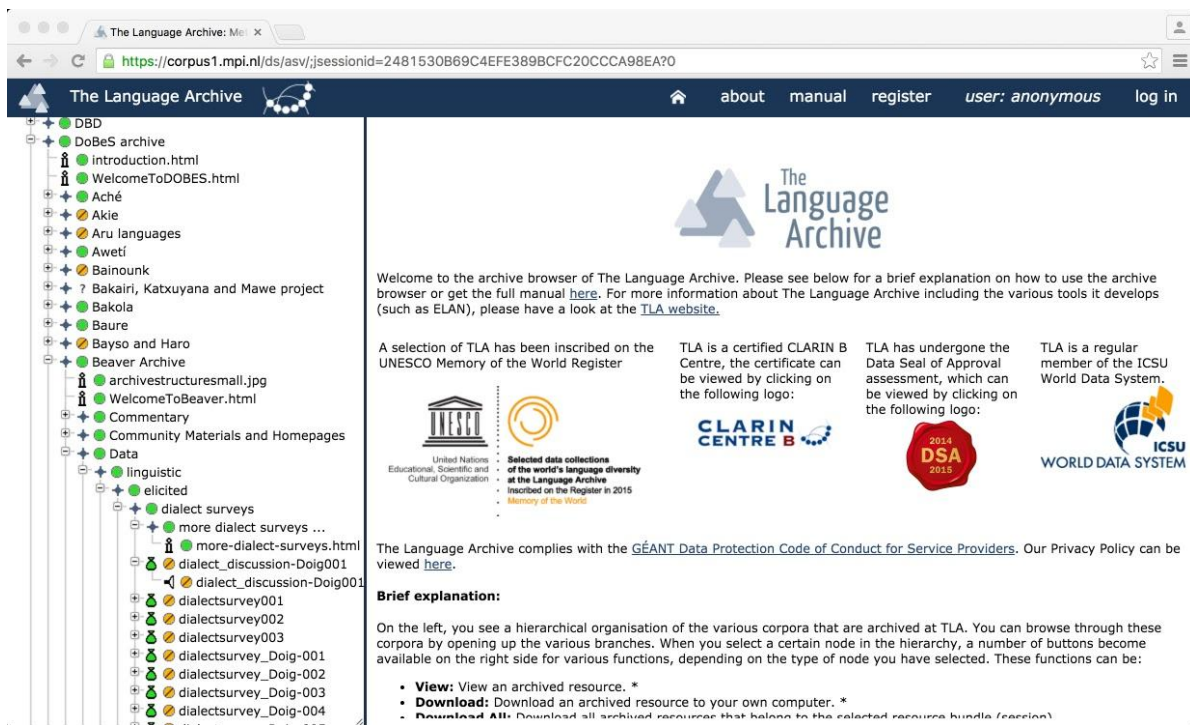
Weitere Differenzierung möglich:

Persona Language Specialist, Persona Typologe, Persona Syntaktiker, Persona Community Member

Survey

- 20 Archive und Portale
- informelle und strukturierte Interviews mit Forschern
- Erfahrung aus dem Betrieb des Language Archive Cologne
- Dialog mit anderen Spracharchiven
- Ergebnisse des Workshops »User Centered Design of Language Archives«

The Language Archive (MPI, Nijmegen)



The screenshot shows the website interface for The Language Archive. On the left is a hierarchical navigation tree with categories like DBD, DoBeS archive, Aché, Aru languages, Awetí, Bainouk, Bakairi, Katxuyana and Mawe project, Bakola, Baure, Bayso and Haro, Beaver Archive, archiviststructuresmall.jpg, WelcomeToBeaver.html, Commentary, Community Materials and Homepages, and Data. Under 'Data', there is a 'linguistic' section with 'elicited' and 'dialect surveys' sub-sections, including 'more dialect surveys ...' and 'more-dialect-surveys.html'. The main content area features the 'The Language Archive' logo, a welcome message, and three certification logos: UNESCO Memory of the World Register, CLARIN CENTRE B, and ICSU WORLD DATA SYSTEM. A red '2014 DSA 2015' seal is also present. Below the logos, there is a privacy policy notice and a 'Brief explanation:' section.

The Language Archive

about manual register user: anonymous log in

Welcome to the archive browser of The Language Archive. Please see below for a brief explanation on how to use the archive browser or get the full manual [here](#). For more information about The Language Archive including the various tools it develops (such as ELAN), please have a look at the [TLA website](#).

A selection of TLA has been inscribed on the UNESCO Memory of the World Register

TLA is a certified CLARIN B Centre, the certificate can be viewed by clicking on the following logo:

TLA has undergone the Data Seal of Approval assessment, which can be viewed by clicking on the following logo:

TLA is a regular member of the ICSU World Data System.

The Language Archive complies with the [GÉANT Data Protection Code of Conduct for Service Providers](#). Our Privacy Policy can be viewed [here](#).

Brief explanation:

On the left, you see a hierarchical organisation of the various corpora that are archived at TLA. You can browse through these corpora by opening up the various branches. When you select a certain node in the hierarchy, a number of buttons become available on the right side for various functions, depending on the type of node you have selected. These functions can be:

- **View:** View an archived resource. *
- **Download:** Download an archived resource to your own computer. *
- **Download All:** Download all archived resources that belong to the selected resource bundle (session).

The Language Archive (MPI, Nijmegen)

The screenshot displays the MPI-ANNEX Interface in a Mozilla Firefox browser window. The interface is divided into several sections:

- Text:** A sidebar on the left with options for Text, Grid, Subtitle, and Timeline (selected).
- Video display:** A central video player showing a scene with several people. Below the video are playback controls (play, stop, previous, next, full screen) and a 'Settings' button.
- Media information:** A panel on the right showing details for the resource 'MISE-1.eaf', media file 'MISE-1.mp4', and elapsed time '00:01:42:000'. It also displays 'Selected chunk' information (Begin time, End time, Text).
- Mini Data Frame:** A panel on the right with a 'Tier' dropdown set to 'none' and a 'Font size' dropdown set to '14'.
- Timeline:** A large panel at the bottom showing a time axis from 00:01:36:000 to 00:01:44:000. It displays linguistic annotations for different tiers (mf1, mb1, gl1, ft1, mf2, etc.) with text in a non-Latin script. A vertical red line indicates the current playback position at approximately 00:01:42:000.

At the bottom of the interface, there is a status bar that reads 'Transferring data from corpus1.mpi.nl...'.

DoBeS-Portal (MPI, Nijmegen)



DOBeS

DOCUMENTATION OF ENDANGERED LANGUAGES

FEATURED MEDIA



RESEARCH PORTAL



GENERAL INTEREST PORTAL



DEPOSIT YOUR DATA

WELCOME TO THE DOBeS PORTAL

The DOBeS Archive contains language documentation data from a great variety of languages from around the world that are in danger of becoming extinct. This portal gives access to the material in the archive and provides information about the DOBeS endangered languages documentation programme.



[North and Meso-America](#) | [South America](#) | [Eurasia](#) | [Africa](#) | [South East Asia and Oceania](#)

This map displays all languages currently in the DOBeS Archive. By clicking on a location you can go to the documentation project pages.

Bahasa Indonesia
English
Español
Français
Português
Русский

Type your search term here

[Search the DOBeS archive](#)

Show only results that are accessible to me

Access and Registration

[Research Portal](#)

[General Interest Portal](#)

[Deposit your Data](#)

[Documentation Projects](#)

[Research Projects](#)

[DOBeS Programme](#)

[Archive Information](#)



BROWSE THE DOBeS
ARCHIVE

 VolkswagenStiftung

© 2000 - 2015

Endangered Languages Archive (SOAS, London)

The screenshot shows the homepage of the Endangered Languages Archive (ELAR). At the top, there is a dark navigation bar with the ELAR logo and the text "ENDANGERED LANGUAGES ARCHIVE" on the left, and "LOGIN | NEW USER" on the right. Below this is a light grey bar with a menu of links: "Home", "About", "Using ELAR", "Depositing", "Documenting languages", "News & events", and "Contact". To the right of the menu are social media icons for Facebook, Twitter, and Email.

The main content area features the large orange text "ELAR" on the left. To its right, the text reads "PRESERVING AND PUBLISHING OF DOCUMENTATION OF ENDANGERED LANGUAGES". Below "ELAR" is a search section titled "SEARCH BY" with a "LOCATION" label and a world map. Underneath the map is a "TEXT" search box with the placeholder "Enter search text..." and a magnifying glass icon. To the left of the search box is the label "LISTS".

Below the search section, there are two columns of links. The first column, under "LISTS", contains "LANGUAGES", "COUNTRIES", "DOCUMENTERS", and "TITLES". The second column, under "HELP", contains "NEW USERS", "USING THE ARCHIVE", "BECOME A DEPOSITOR", and "LANGUAGE DOCUMENTATION".

The footer is a light grey bar containing three logos: the Hans Rausing Endangered Languages Project logo, the SOAS University of London logo, and the Arcadia logo with the text "Funded by Arcadia".

Endangered Languages Archive (SOAS, London)

Endangered Languages Archive

Not logged in. [Login](#) | [New user](#) | [Search](#) | [Home](#)

Search

Found 90305 resources in the archive (page 1 of 11289)

1 2 3 4 5 6 7 8 9 ... [next >](#) [last >>](#)

Search ELAR

[Reset keywords](#)

Language

- AVSL (84)
- Adelaide dialect (171)
- Ainu (489)
- Asherinka Perene (105)
- Avatime (Dominant) (214)
- Awiakay (84)
- Ayuŋla Mxe (115)
- Bajjika (175)
- Baram (156)
- Bierebo (89)
- Brisbane dialect (163)
- Cappadocian (77)
- Cashibo-Cacatabo (293)
- Chatino, Nopala (73)
- Chatino, Zacatepec (68)
- Chinese (157)
- Choguita Rarámuri (750)
- Chuxnabán Mixe (77)
- Cicpu (259)
- Ctz (208)
- Dalabon (459)
- Dhao (74)
- Duoxu (79)
- Durvud (299)
- English (2236)
- Gija (1095)
- Goemai (79)
- Great Andamanese (291)
- Guari-Guari (80)
- Gurene (279)
- Gurindji (114)
- Gújjolsay Eegmaa (95)
- Hahon (77)
- Hindí (482)

- Elizabeth Marrkilyi Ellis**
Australian National University and Ngaanyatjarra Community Member
- Pierpaolo di Carlo**
Linguistics Department, University at Buffalo, State University of New York USA
Associated deposits:
Verbal art, ethnohistorical performances and welcome rituals from Ngun and Ajumbu
- Verbal art, ethnohistorical performances and welcome rituals from Ngun and Ajumbu**
Pierpaolo di Carlo
... documenting art, performances and rituals of the Ngun and Ajumbu communities within the language ecology of Lower Fungom (NW Cameroon)
- IHUOMA AKINREMI**
UNIVERSITY OF JOS

How to use search

You can search in two ways:

- enter text in the search box and press 'Search'. Search is not case sensitive, and variations of words are found, e.g. 'Village' finds 'villages' and 'Indian' finds 'India'; or
- click a keyword in the left panel to find a set of resources. Click another keyword to refine the results (a black keyword) or to find a new set (a brown keyword)

To refine your search:

- enter two or more words for results containing all those words; e.g. entering 'nigeria' and 'audio' finds the deposit *Damakawa worl/sit* which includes recordings made in northern Nigeria.
- use the keywords in the left panel to browse and select further categories; e.g. if you search for 'nigeria' and 'audio', a list (under 'tags') includes place and language names: Akoko, Ikaann, Damakawa and Sakaba. Click one to find a resource pertaining to that name

To reset search and display all keywords, press 'Reset keywords'.

Colour coding of results

Search results can include deposits, bundles (file groups within deposits) and people. These are colour coded:

- A** deposit: click on its title to view the deposit
- B** bundle (group of files within a deposit): click on the icon or title to see the files and more metadata
- C** a person (usually, a depositor): click on the name to see more details, or click on one of the associated deposits.

ELDP (SOAS, London)



California Language Archive (UC Berkeley, USA)

CALIFORNIA LANGUAGE ARCHIVE

A catalog and online archive of the indigenous languages of California, western North America, and the Americas

Depositor login

HOME EXPLORE COLLECTIONS ABOUT US ACCESS HELP

Search collections


 Include material only if it contains online content

Contributing repositories:
Bancroft Library | Berkeley Language Center
Phoebe A. Hearst Museum of Anthropology
Survey of California and Other Indian Languages

Featured: Alice Shepherd Papers

The Papers document the research of Alice Shepherd (née Schlichter) on the Yukian languages, primarily Yuki proper. The collection includes Shepherd's research notes, manuscripts, theses, and correspondence. It also includes Yuki, Huchnom, and Coast Yuki vocabulary. [\[Learn more.\]](#)

Explore collections



© 2015 The Regents of the University of California. The California Language Archive is managed by the Survey of California and Other Indian Languages, Department of Linguistics, UC Berkeley

California Language Archive (UC Berkeley, USA)

CALIFORNIA LANGUAGE ARCHIVE

A catalog and online archive of the indigenous languages of California, western North America, and the Americas

Depositor login

[HOME](#) [EXPLORE COLLECTIONS](#) [ABOUT US](#) [ACCESS](#) [HELP](#)

GO

Item description

Handgame song

Your search: LA 51.001

Item number: LA 51.001

Contributors: Ray Rackley (researcher), Eva Thompson (consultant)

Language: Panamint

Duration: 1 minute, 7 seconds

Catalog history: Digital asset LA51.001.001.wav was formerly segment number 001_1.

Location: Lone Pine

Collection: The Ray Rackley collection of Panamint sound recordings


Repository: Berkeley Language Center

Digital assets:

1. LA51.001.001.wav [Download MP3]
Play MP3

© 2015 The Regents of the University of California. The California Language Archive is managed by the Survey of California and Other Indian Languages, Department of Linguistics, UC Berkeley

AILLA (University of Texas Austin)



The Archive of the Indigenous Languages of Latin America

Languages
English

Register/Login

Welcome
- Mission
- History
- People
- Collections

How to Use

Browse

Depositors

Rights

Languages

Events

AILLA papers

Contact

Help

Links


Contribute


Welcome to AILLA

AILLA is a digital archive of recordings and texts in and about the indigenous languages of Latin America.

Access to AILLA and its resources is always free of charge. Most of the resources in the collection are available to the public, but some have special [access restrictions](#).

You must [Register and Login](#) to open or download any media file.

 **Users:** start by [browsing the catalog](#) or use the navigation menu on the left to find more information about the archive and the indigenous languages of Latin America.

 **Depositors:** If you want to deposit materials in AILLA, [start here](#).


The latest [News and Announcements](#) about AILLA can be found [here](#).

The Collection

The heart of the collection is recordings, both audio and video, in a wide range of genres: narratives, chants, oratory, conversations, songs, and more. Many recordings are transcribed and translated into Spanish, English or Portuguese.

The archive also contains a wealth of language documentation materials: grammars, dictionaries, ethnographies, and field notes. The collection includes teaching materials for bilingual education and language revitalization programs.

We are committed to serving the indigenous people of Latin America and the scholars who study their languages. We are always interested in hearing from you. Please [contact us](#) if you have any questions, comments, or suggestions.



AILLA is a joint effort of the LILAS Benson Latin American Studies and Collections, the Department of Linguistics, and the Digital Library Services Division of the University Libraries at the University of Texas at Austin.
AILLA is also grateful for support from the National Endowment for the Humanities and the National Science Foundation.
[Legal Disclaimers](#) | [Graphics sources](#)

AILLA (University of Texas Austin)



The Archive of the Indigenous Languages of Latin America

Languages
English

Register/Login

Welcome

How to Use

Browse

- By Collection
- By Language
- By Lg Code
- By Country
- By Depositor
- Keyword Search

Depositors

Rights

Languages

Events

AILLA papers

Contact

Browse by Collection

Aguilera, Oscar
Amith, Jonathan
Amith, Jonathan
Amith, Jonathan
Amith, Jonathan
Amith, Jonathan
Basso, Ellen
Beam de Azcona, Rosemary
Beier, Christine
Beier, Christine
Bermúdez, Natalia
Blunk-Fernández, William
Bohnemeyer, Juergen
Bolaños Quiñonez, Katherine
Boudreault, Lynda
Bricker, Victoria
Broadwell, George Aaron
Brody, Jill
Brody, Michal
Bucca, Salvador
Cairns, Charles
Campbell, Lyle
Carrasco, Michael
Castillo García, Rey
Cervantes, Laura
Chanin, Mac

The Chilean Languages Collection
Balsas Valley Nahuatl, Central Guerrero, Mexico
Yoloxóchtli Mixtec Language Documentation Project
Nahuatl Documentary "Silvestre Pantaleón"
Nahuatl of the Sierra Nororiental de Puebla, Mexico
Kalapalo Collection
Southern Zapotec Languages Collection
Nanti Collection
Matsigenka Texts
Miskitu Language Collection
Yucatec Maya Collection
Mesospace Collection
Colombian Languages Collection
Sierra Popoluca Collection
Mayan Languages Collection
MesoAmerican Language Collection
Tojolab'al Collection
Yucateco Collection
Argentinian Languages Collection
Southern Paiute Collection
MesoAmerican Languages Collection
Yucatec Maya Collection
Yoloxóchtli Mixtec Language Documentation Project
Indigenous Languages of Costa Rica Collection
Kuna Collection

Zusammenfassung

- Archive sind meist entweder Produzenten oder Konsumenten orientiert
- Existierende Spracharchive unterscheiden sich stark in der Ausrichtung
- Existierende Spracharchive unterscheiden sich stark im Funktionsumfang
- Anspruch: Workspace, Archiv, Publikationsplattform
- Stakeholder formen die User Experience von Spracharchiven

Ausblick

- User Experience/Interaktion hierauf aufbauend gestalten
- Die Rolle als Vermittler zwischen Produzenten und Konsumenten ernst nehmen
- Zwischen Nutzungsszenarien differenzieren
- Priorisierung von Kernfunktionalitäten gegenüber Mehrwertdiensten
- Qualität von Daten und Metadaten
- Hürden für Zugänglichkeit abbauen

Workshop “Open Access und Open Data in Language Research and Documentation: Opportunities and Challenges”, Universität zu Köln, 10.-12. Oktober 2016

Danke!

Slides: tr.im/dhd2016

Twitter: [@LACologne](https://twitter.com/LACologne)

Jonathan Blumtritt: jonathan.blumtritt@uni-koeln.de

Felix Rau: f.rau@uni-koeln.de

KA³



GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung