the ‘reach’ of digital language archives: towards criteria for evaluation

David Nathan

Endangered Languages Archive
SOAS, University of London
www.elar-archive.org

Paradisec RRR 2013
University of Melbourne
2 December 2013
Aims

- initial suggestions towards a set of criteria for considering and evaluating the “reach” (access, accessibility, awareness) of our archives
- complementary to other crucial areas such as preservation etc
- approaches to “reach” could be much more scientific than they are
Archive evaluation schemes

- Open Archival Information Systems OAIS (2002)
  National Initiative for a Networked Cultural Heritage
- Data Seal of Approval (2009)
- TAPS (2010)
OAIS reference model

- 3 ‘packages’ ingestion, archive, dissemination
- fairly one way flow
- recognise communities of producers and users
Archive evaluation schemes

- Open Archival Information Systems OAIS (2002)
  National Initiative for a Networked Cultural Heritage
- Data Seal of Approval (2009)
- TAPS (2010)
TAPS

- TAPS (Target, Access, Preservation, and Sustainability, Chang 2010) addresses:
  - recommended best practices for the long-term preservation of digital information (preservation)
  - areas of special concern to linguists and language communities (access)
TAPS on “reach”

- Will the audience (designated communities) that I wish to reach be able to access archive materials?
- Will members of the be expected to have access to the Internet?
- Will they need to maintain an email address?
- Will the metadata be available in English only, or will it be available in another language that is more accessible to them?
- Will the archive charge fees for copies of data on media that are usable by them?
“Reach” 1

- *acquisition*: policy and needs-driven acquisition, ingest and curation, increases coverage
“Reach” 2

- *audiences*: understanding of audiences to provide appropriate services for them, e.g. their languages of access, their varied technological and information literacies, interface design and usability
Coección: Southern Zapotec Languages Collection

Visualizar los recursos en esta colección

<table>
<thead>
<tr>
<th>Título</th>
<th>Colección de Lenguas Zapotecas de la Sierra Sur</th>
</tr>
</thead>
<tbody>
<tr>
<td>Título inglés</td>
<td>Southern Zapotec Languages Collection</td>
</tr>
<tr>
<td>Lenguas en esta colección</td>
<td>Zapotec, Coatec, Zapotec, Coatec, Zapotec, Isthmus, Zapotec, Miahuatec, Zapotec, Sierra Juárez</td>
</tr>
<tr>
<td>Colecionista(s)</td>
<td>Rosemary Beam de Azcona</td>
</tr>
<tr>
<td>Depositante(s)</td>
<td>Rosemary Beam de Azcona</td>
</tr>
<tr>
<td>Sitio del Proyecto/Coleccionista</td>
<td></td>
</tr>
<tr>
<td>Descripción</td>
<td>Esta colección incluye grabaciones de audio en varias variedades de zapoteco habladas en Oaxaca, México y textos manuscritos. Las grabaciones representan una gran variedad de géneros de discurso. Los textos incluyen notas del campo, juegos de datos, y transcripciones y traducciones de textos grabados.</td>
</tr>
</tbody>
</table>

Referencias

Resumen de los contenidos de la colección

<table>
<thead>
<tr>
<th>Géneros</th>
<th>Lista de palabras; Encuentro; Instrucciones; Historia; Juego de datos; Entrevista; Gramática; Léxico; Narrativa; Descripción; Comentario; Esbozo; Procedimiento; Conversación; Juego de habla</th>
</tr>
</thead>
<tbody>
<tr>
<td>Número de archivos archivales</td>
<td>1046</td>
</tr>
<tr>
<td>Porcentaje de archivos restringidos</td>
<td>14</td>
</tr>
<tr>
<td>Número de grabaciones</td>
<td>46</td>
</tr>
<tr>
<td>Duración total de audio</td>
<td>16:37:67</td>
</tr>
</tbody>
</table>
Reach
 delivery: resources available to suit users' preferences eg download, view-in-browser, through apps or other means
 access management: follows depositors' and communities' preferences, users have ways of navigating through collections of varied accessibility and applying for and negotiating for access
 information accessibility: desired content is accessible to users (contextualisation, complexity, structure, language, modality)
Welcome to the catalog of the PARADISEC collection
“Reach” 3

- *discovery*: drawing on understandings of audiences in order to help them browse, navigate, search, identify and select their items of interest
Documentation of Mavea

Contributor
Valerie Guerin (162)
Gabriel Torno (2)

Participants
Sera Lima Lowet (88)
Allan Natu Lowet (18)
Elsie Fopua Kaman (15)
Mosela Vomei Kaman (10)
Valerie Guerin (5)
Fred Kaman (4)
Jo Tavon Livo (4)
PFL (4)
Paul Sope Livo (4)
James Sesei Livo (3)
Lowet Daldal Morris (3)
Pupu Moldovo Morris (3)
Rogen Molavea Lowet (3)
Gabriel Torno (2)
John Molsi Livo (2)
Judy Vokarae Livo (2)
Morris Tov'aoi Kaman (2)
Peter Vupoaiai Lowet (2)
Alfred Moltak Kaman (1)
Johnatan Nono Livo (1)
Jona Parparu Morris (1)
Lina Vatari Simpia (1)
Rolin Volfi (1)

Deposter
Valerie Guerin
Nationality: French
Affiliation: University of Hawai'i Manoa

Your access
Default access protocol: U
Your access roles:

Deposit
Group represented: Mavea speaking community, located on Mavea Island, and Deproma, Espiritu Santo Island, Vanuatu.
Location: Recordings were all created in Vanuatu. Locations include: Vunopuna, Saoori, and Vunepua (on Mavea Island), Deproma (Espiritu Santo Island), Port Vila (Efate Island), Aore and Tanna Islands.
“Reach” 4

- *delivery*: making available selected resources according to users’ preferences whether by download, view-in-browser, through apps or other means; also considering emerging devices
Choguita Rarámuri description and documentation

Found 1 bundle in this deposit (page 1 of 1)

batusia chukú

batusia_chuku.mpg

Access protocol: U R C S

ID: batusia_chuku
Title: batusia chukú
Date created: 10/9/07
Description: Video recording of MDH preparing the dough for tortillas

Keywords: Choguita Rarámuri - Conversation - Procedure - MDH
anjinmul (AHN-jin-mool)
your nose

-rra - /rr/ i.e. return; go back; come back (depending on choice of directional prefix). Artirra nangaj. He came back yesterday. Yartiirra nangaj. He went back yesterday. Barda yangmanawulakun yangmanirra nyurraang alka. Then you go down back to your place in the sea.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Stem</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>ngarrir</td>
<td>ngar</td>
<td>singular</td>
</tr>
<tr>
<td>ngarriri</td>
<td>ngari</td>
<td>head</td>
</tr>
<tr>
<td>ngarruru</td>
<td>nguru</td>
<td>hand</td>
</tr>
<tr>
<td>ngarru</td>
<td>nguri</td>
<td>hand/finger</td>
</tr>
<tr>
<td>nangad</td>
<td>nangad</td>
<td>teeth</td>
</tr>
<tr>
<td>nangad</td>
<td>nangad</td>
<td>singular</td>
</tr>
</tbody>
</table>

Aboriginal person

Acanthopagrus berda Acanthopagrus berda

Acanthopagrus dussumieri Acanthopagrus dussumieri

about concerning

key

Acacia leptocarpa wattle

manjinaajini

nuruk

nuruk
“Reach” 5

- **access management:**
  - resource delivery follows depositors’ and communities’ preferences
  - access conditions and actions are transparent
  - users have ways of applying for and negotiating for access (where possible)
Access protocol

URCS (155)
URCS (3)

Language
Kubokota (148)
Luqa (10)

Type
Audio (151)
Document (4)
ELAN (98)
Image (11)
Text (93)

Genre
Custom description (13)
Custom narrative (16)
Custom story (1)
Directional story (4)
Elicitation (17)

Topic
Abandonment (1)
Adultery (1)
Artifacts (1)

Deposit contents
The deposit comprises 150 audio files, as well as written texts, elicitation materials and participant observation notes.

Genres include traditional narratives, procedural and route descriptions, personal stories and accounts of everyday events in which the researcher participated. Elicited materials include responses to the caused positions and
In order to be considered for access, you will need access rights. You will receive an email that even if your request is approved, you will not be able to view this particular file. Apply for access.
Concerning Edward Garrett's application

Tell the depositor who you are and why you would like access to restricted material in the deposit.

Hi Anthony, I'd like to hear the recording of the 'well' conversation.

Send request

User
Access levels

We recommend that you make publicly accessible any materials that are not sensitive by their nature; that is, that are not sacred or dangerous or embarrassing for the speakers. This is best over the long term for speakers and researchers alike. If you wish to discuss your particular circumstances, please contact us.

**Level 1. Public access**

Users have full access to these materials after agreeing to our Terms and Conditions and logging in. Choose this level for all non-sensitive materials. This is the default level.

**Level 2. Password**

You define a password with an optional hint. You can provide access to others by sharing the password with them. Passwords with hints allow a set of people to share resources easily. Hints can be something like that everyone in your in-group knows, but no outsider would know. Example: password = 7awin7; hint = 'zoque amigo'. You can use one password + hint for one set of resources and another for a different set. This is the most flexible option.

**Level 3. Time limit**

After the date specified, access will change to Level 1. Example: 2050-01-01 (January 1, 2050) This option is good for resources that are only sensitive for a period of time, for example, the lifetime of the narrator or the five years until you finish your thesis. You will still be able to access these materials yourself with the password.

**Level 4. Depositor control**

Users must contact you directly to ask you for the password. AILLA will provide contact information. Note that you must maintain your contact information and be prepared to respond to requests from people you do not know.
“Reach” 6

- *information accessibility*: the actual desired content is accessible to users, whether in terms of contextualisation or appropriate complexity, language, or modality
This (lady) is the mother of my grandmother

كانَتُ هذِهُ هِيَ باعتِبارِ إمَّ الْحَوْشِ شَيْ إِذْ لَيْدُي أَذْ خُوَسُ

In the winter

وَكَانَ بَالْحُامِٰدِ

they used to make for her, a 'Kursi Jafuf'

يَسوُىُ هَا كَرْسِيْ حَطُوف

Eli Timan’s Jews of Iraq website: http://www.jewsofiraq.com/texts/kersijafuf001.xml
Try repositioning and resizing the bubbles. Click on the bubble to cycle through different bubble styles. Then click on the button to show the new CSS settings.

Speech bubble demo: http://lah.soas.ac.uk/projects/dev/bubble-player/wilbur.html
“Reach” 7

- **promotion**: depositors, funders, archives and others can promote awareness
  - Wilbur (LDD12): promotion through local archiving
  - Johnston & Schembri: promotion through teaching and social media - 95 ELAR users, of which 2/3 registered in last 12 months
Session description: The Morehead culture is unique in the world in using a senary (base-six) number system. This system is traditionally employed in yam counting, which is an important ritual in the intense exchange relations between villages. These videos are from a yam counting procedure that took place in preparation of a wedding. Many unnamed participants were involved.

tci20100901-01CDv-01 = carrying the yams and preparing for counting;
tci20100901-01CDv-02 = counting, tci20100901-01CDv-03 = final count of the tarumbas
Keywords: yam-counting, procedural, ceremony, marriage (IMDI archive)
“Reach” 8

- *communication ecology*: archive dissemination can be complemented by
  - other web sites (advice, activities etc)
  - Facebook pages (seeking contact, resources etc)
  - blogs
  - talks, publications
  - training
  - personal networks
Imagine ... a world without PARADISEC

May 15, 2013, 1:21 am by Jane Simpson

Imagine ... a world without memories is the evocative and chilling title of a project organised by the National Committee of Australia for the UNESCO Memory of the World.
The West Arnhem Land song project – podcast with Linda Barwick

This week in the world of ELAR podcasts, Linda Barwick tells us about her joint documentation project with Allan Maret and other linguists and musicologists in West Arnhem Land, Australia. Linda talks about how the project came out of an “interest in how the musical traditions [of West Arnhem Land] relate to language and how music communities navigate around multilingual layers of affiliation and communication.” Have a listen to the podcast at SOAS radio: Western Arnhem Land Song Project.

An overview of the documentation project can be found on its ELAR deposit page.
“Reach” 9

- *feedback channels*: users can provide feedback to depositors or enhance deposits with user-generated content; archive can provide usage information to depositors (and communities)
Measuring impact?

- member of OLAC and DELAMAN
- 140+ deposits, 81,000 files, 35,000 bundles, 10TB
- media (sound, video) – 47,000 recordings, 6,500 hours
- online about 700 unique visitors/day, 130,000 page views per month
- registered membership > 1100
  - > 10% community members
- but we also need to work out ways to detect and report the value of usage of our archives
Distribution of ELAR users (n=1040)

- Researchers: 74%
- Speakers: 10%
- Depositors: 13%
- Journalists/activists: 2%
- Artists/creative professionals: 1%

ELAR users tell us …

- “I study the syntax of wh-constructions in Cape Verdean Creole and I'm interested in Portuguese contact-language situations”
“Value statements”

“…I am a teacher of the Deaf working in a Bilingual Program that uses Auslan. Use of the corpus will enable teachers and students at TPS to explore their own language and its features using real examples from a diverse range of Auslan users and range of texts”

“…I am a linguist and language teacher, I work with the Northern Arapaho language, teaching the language at the University of Wyoming, and creating instructional materials for use in all Arapaho language classrooms.”
“Value statements”

- “Kuuku Ya'u descent”
- “member of aboriginal corporation as a committee member trying to connect to my heritage through the research of our song and dance ceremonies”
- “I am a member of the Unga tribal village and doing research for a project relating to my heritage”
- “My son has married a goemai woman. I'm interested of this language”
Conclusion

- suggested criteria for evaluating output, impact and value of digital language archives:
  - acquisition
  - (understanding) audiences
  - discovery
  - delivery
  - access management
  - (information) accessibility
  - promotion
  - communication ecology
  - feedback
References