Scaling access management: an endangered languages archive perspective

David Nathan
Kakia Chatsiou

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

second innet conference, 6-7 Sept, 2013
Scaling access … summary

- outline problem as we see it
- access as users’ process of ‘getting’ materials
- other forms of access enhancement
  - (ELAR has implemented only some of these)
- possible implications for broader infrastructure
Archives, access and ambiguity

- What do we mean by
  - 'archive'?  
  - 'access'?
About ELAR

- a digital endangered languages archive
  - policy development
  - preservation
  - cataloguing and dissemination
  - training and advice
  - materials development and publishing
- since 2005
- at SOAS University of London
About ELAR

- 140+ deposits, 10TB and growing
  - media (sound, video) – 30,000 recordings, 6,500 hours
  - graphics (images, scans)
  - structured data (annotations, transcriptions, databases, lexica)
  - text (fieldnotes, grammars, description, analysis)
  - metadata (structured, standardised contextual information about the materials)
  - transcoded versions of files into preservation and dissemination friendly formats
General approaches to access

- declarative statements about audiences served, “three major user groups can be identified: The speaker community, the scientific community, i.e. linguists and scholars of related disciplines, and the general public” (Schwiertz)
  - are we really catering for them? What do we really know about their goals, skills etc?
  - moderation of potential contributions from various audiences?
  - ‘voice’ – how are contributions from various audiences framed/presented in the archive?
General approaches to access

- by declarative statements about respect of sensitivities and thus distribution (‘access protocol’)

“By default, all archived material except … the metadata descriptions are not openly available and access therefore will be restricted” (Wittenburg 2005)
General approaches to access

- by comprehensive methods for metadata-led resource discovery
  - tends to be discovery of *existence*
  - tends not to extend to ease of access, negotiation of rights, level of access to content
Discussion

- we find little primary research or data about who users are, what they need and how they interact
- access has many components, not just open vs closed or managed availability
  - interface and navigation/interaction design
  - processes of negotiation
  - format, rendering, selection, version, or bandwidth of supplied resource
  - results/outcomes/value of the process
Discussion

- the access process *changes*:
  - conventions and device technologies change over time
  - norms, expectations and skills change with types of users
- we may not be very scientific or transparent about our assumptions, eg
  - who really are the users; what do they do
  - cost/effort caused to user by ‘friction’ of a managed access process
Discussion

- archives need to understand the *value* of an access to a user, eg.
  - UKDA asks users to associate a download with a use description (ie say why they are downloading)
  - ELAR asks users why they are registering
  - Woodbury, Austin: we don’t seem to have particularly high levels of usage (downloads etc)
- yet, our impact is significantly evaluated by usage factors
“Value statements”?

- I found your site while looking up my grandmother, and I found her on your site speaking our language. And I would love for my children her great grandchildren to hear our language coming from her.
- I'm a native zapotec speaker and a linguist, currently I'm working in order to document my own language.
- My son has married a goemai woman. I'm interested of this language.
- I used to be a Pow Wow singer in Oregon. We sang mostly Lakota and Ponca songs.
“Value statements”?

- I am a linguist working on an endangered language (Siouan, North America). I've been doing fieldwork there for the past 10 years and have gained a lot of experience in language documentation as well as revitalization. I …would like to use archived materials to conduct a study on complex attribution.

- I'm a native zapotec speaker and a linguist, currently I'm working in order to document my own language.
“Value statements”?

- I found your site while looking up my grandmother, and I found her on your site speaking our language. And I would love for my children her great grandchildren to hear our language coming from her.
- My son has married a goemai woman. I'm interested of this language.
- I used to be a Pow Wow singer in Oregon. We sang mostly Lakota and Ponca songs. I am interested in learning more about the local languages, especially Warm Springs languages.
Distribution of ELAR users (n=1300)

- Researchers: 77%
- Depositors: 13%
- Journalists/activists: 2%
- Artists/creative professionals: 1%
- Speakers: 7%
Scaling responses?

- the technical and engineering responses mainly target
  - upload tools and ‘automation’
  - discovery enhancement (reducing burden of discovery)
Elephant in the room

- how to scale managed access since
  - there is consensus on respecting access protocols
  - one resource potentially has many accesses
  - language documentations often consist of 1000s of files
  - facilitated discovery amplifies the importance of scaling managed access
Solutions?

- too hard! – ignore
- depositors supply fixed preformulated values for open or closed access
- increase value of resources so that cost/benefit ratio improves for users – publishers’ model
Solution?

- social networking
  - ELAR’s subscription model/system
URCS

- **U** – resource available to *all* registered users
- **R** – resource available to users registered as *researchers*
- **C** – resource available to users endorsed as members of relevant *language community*
- **S** – resource available to users who have been given *individual* access rights for that resource
Documention of Ratahan

Found 1 bundle in this deposit with keyword URCS (page 1 of 1)

RTH067_welli

RTH067_welli.wav Access protocol: URCS
This file is not available to all users. In order to be considered for access, you must apply to the depositor for individual access rights. You will receive an email with the outcome of your request. Note that even if your request is approved, you still may not be allowed to view this particular file. Apply for access.

RTH067_welli-MPEG-1 ELAN.mpg Access protocol: URCS

User
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
Depositor
Concerning Edward Garrett’s application

Give a reason. This will be shared with the user, and may also be used by ELAR to improve its services.

I'm happy to accept your test application.

Approve application

Depositor
Depositor
User
User
What do requesters say?

- “I have just moved with my wife to Madras and would like to learn some of the Kiksht language to show respect when interacting with members who speak the language.”

- “I am at SOAS looking at their access system, so much better than AIATSIS :-). I was wondering if i can access just one of the Mawng songs from the Manmurulus. Cheers“
What do requesters say?

- “My son has married a goemai woman. I'm interested of goemai's culture and language”

- “… I am currently teaching the courses 'Mesoamerican languages A' and 'Language I' and will dedicate a lecture to language death and revitalization. Therefore, it would be of great value to me to get access to some of your material for teaching purposes - I would like to introduce the students to your documentation project and ELAR in general.”
What do requesters say?

- “Hi there, looking to hear some old Golpa being spoken in order to assess differences in intonation and emphasis for my work on the Yan-nhangu dictionary”

- “I research the importance of bees in human culture around the world. I am hoping to learn about Reyesano knowledge of bumble bees. I hope you can help. Thank you for your time”
Social networking as a scaling solution

- scalable - distributes effort
- … to the very stakeholders in the transaction
- sharing of additional information
  - home page, and communications, provide rationale and detail for restrictions
  - users and depositors share advice, data, connections, arrange further collaboration
- development of relationships, community
- value traces result from usages - archive can ‘learn’ dynamics and improve
Social networking as a scaling solution

- build and demonstrate trust by providing users and depositors with an *access monitoring model*
  - reports to depositors on who downloaded what and when
- archive takes appropriate role as *facilitator* of access processes and relationships
- access (as supply) understood as a situated transaction
- automated and scalable
Other aspects of access

- we see several types of access as best taking place at the presentation interface
  - lowers barriers to participation
  - few special tools or skills needed (‘if you can see it, you can interact with it’)
  - actions/contributions are contextualised
  - results are tangible
  - - as in WYSIWYG software that most users are familiar with
Interface-located functions

- edit deposit’s ‘home page’
### Choguita Raramuri description and documentation

**Language:** Choguita Raramuri
- **TAR**

**Depositor:** Gabriela Caballero

**Location:** Mexico

**Summary of deposit**
This deposit consists of audio and video recordings with transcriptions and photos of speakers of Choguita Raramuri, an underdescribed language spoken in the state of Chihuahua, Mexico, resulting from fieldwork conducted between January 2006 and December 2009.

**Resumen del depósito**
Este depósito consiste en grabaciones de audio y video con transcripciones y fotos de hablantes de Choguita Raramuri, un idioma sin descripción adecuada que se habla en el estado de Chihuahua, México, que resultan de trabajo de campo realizado entre enero 2006 y diciembre 2009.

**Group represented**
Raramuri from Choguita, Guachochi municipality

**Language information**
Choguita Raramuri (Tarahumara)

See also Gabriela Caballero's dissertation: *Choguita Raramuri (Tarahumara) phonology and morphology* © 2008

**Language**

**Group represented:**
This information appears in the right panel of your deposit home page.
Enter the name of the cultural or linguistic group(s) represented by your documentation.

Rarámuri from Choguita, Suárez municipality

**Language name:**

<table>
<thead>
<tr>
<th>Order</th>
<th>Language name</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Choguita Rarámuri</td>
</tr>
</tbody>
</table>

Enter the subject language name as you would like it to appear in the catalogue.

Add another item

**Language code:**
If known, please enter a comma-separated list of (ISO 639-3) language code(s) characterizing the content of this deposit. Consult the Ethnologue for a listing of language codes.

TAR

**Language information:**
This information appears in the right panel of your deposit home page.
Enter any additional information about the documented language(s), including any other names used for the language.

Choguita Rarámuri (Tarahumara)
See also Gabriela Caballer's dissertation: [link](http://linguistics.berkeley.edu/dissertations/Caballero_dissertation_2008.pdf)
Choguita Rarámuri (Tarahumara) phonology and morphology 2008

**Depositor**
Deposit

Catalogue description:

... audio and video recordings with transcriptions and photos of speakers from Chihuahua, Mexico. Includes myth and historical narratives, oratory, interviews, conversations, activities, ritual song and prayer, and language teaching.

Summary of deposit:

This information appears in the Home tab (centre panel) of your deposit home page.

Enter a summary of the documentation.

This deposit consists of audio and video recordings with transcriptions and photos of speakers of Choguita Raramuri, an undescribed language spoken in the state of Chihuahua, Mexico, resulting from fieldwork conducted between January 2006 and December 2009.

Deposit contents:

This information appears in the Home tab (centre panel) of your deposit home page.

Describe the contents of the documentation.

The deposit comprises:

- over 750 audio and video recordings of texts, elicitation sessions and community events;
- ELAN and Transcripter transcription files;
- Word files with transcriptions, translations, analyses and annotations;
- Excel databases of lexical items and prosodic and morphological properties;
- still pictures;
- metadata including descriptions of documents, recording

Special characteristics:

This information appears in the Home tab (centre panel) of your deposit home page.

Describe any unique or outstanding features of the documentation.

This deposit includes documents created by community members.
Country: Mexico

Current coordinates

Latitude: 27.000408
Longitude: -106.907225
Source: User-submitted
Interface-located functions

- edit ‘bundle’ (including access)
Documentation of Reyesano

Language: Reyesano [SAI]
Depositor: Antoine Guillaume
Location: Bolivia

Summary of deposit
Audio and video recordings with transcriptions and photos of speakers of Reyesano, a highly endangered language spoken in Reyes, Bolivia, resulting from fieldwork conducted between April 2003 and October 2005.

Resumen del depósito
Este depósito consiste en grabaciones de audio y video con transcripciones y fotos de hablantes de Reyesano, un idioma en mucho peligro de extinción que se habla en Reyes, Bolivia, que resultan de trabajo de campo realizado entre abril 2003 y octubre 2005.

Deposit contents
The deposit comprises over 230 audio and video recordings, transcriptions and Spanish and English translations of some of the recordings, photos, dictionaries and grammars.

Contenido del depósito
El depósito consta de más de 230 grabaciones de audio y video, transcripciones y traducciones en castellano e inglés de algunas grabaciones, fotos, diccionarios y
Documentation of Reyesano

Found 245 bundles in this deposit (page 1 of 31)
1 2 3 4 5 6 7 8 9 ... next > last >

Abrazar [8 August 2012, 6:02 pm]
Personal history
Recorded on 24 May 04
Personal story. Encounter with a jaguar during a fishing expedition

Edit bundle
Hide bundle
Keywords: Reyesano - Personal history - fishing expedition - jaguar - Gurumercindo - Yanana

Adelanto [8 August 2012, 6:02 pm]
Interview
Recorded on 13 July 2005
Antoine interviews Teófilo who talks (negatively) about the life in Reyes, about how poor they are, how nothing happens. Teófilo also talks about the path to Rurrenabaque when there was no highway.

Edit bundle
Hide bundle
Keywords: Reyesano - Interview - life in Reyes - path to Rurrenabaque - poverty - Antoine Guillaume - Teófilo Mano

Alforjas [8 August 2012, 6:02 pm]
discussion
Recorded on 13 July 2005
Berna talks about traditional bags and clothing that her ancestors’ grandmothers would weave

Edit bundle
Hide bundle
**Basic information**

**Identifier:**
A short identifier will help you and others to locate specific resources.
Enter an identifier for this bundle.
Abrazar_ab

**Primary title:**
Use this field as a public-facing reference to this bundle.
Enter a short, descriptive title for this bundle.
Abrazar

**Alternative title:**
Consider using this field for target language titles.
Enter an alternative title for this bundle.

**Description:**
Enter a long description of this bundle.
Personal story. Encounter with a jaguar during a fishing expedition
**Abrazar**

**Keywords**

**Language tags:**
Reyesano

Language tags show up as navigational filters on the left side of your deposit home page. Note that it may take up to a day before changes show up in the interface.

Enter one or more language names to categorise this bundle. Where reasonable, use names that others have already used.

**Tags:**

Free tags help you to use open-ended vocabularies to categorise your resource bundles. Note that it may take up to a day before changes show up in the interface.

Enter one or more tags which categorise this bundle.

**Genre tags:**

Personal history

Genre tags show up as navigational filters on the left side of your deposit home page. Note that it may take up to a day before changes show up in the interface.

Enter one or more tags to categorise this bundle by genre. Where reasonable, use tags that others have already used.

**Topic tags:**

jaguar  fishing expedition

Edit bundle  Hide bundle
Participant tags:
Gumerando Ymana

Participant tags show up as navigational filters on the left side of your deposit home page. Note that it may take up to a day before changes show up in the interface.

Enter one or more tags to categorise this bundle by participant.

Session information

Date created:
24 May 04

Devices:

Add another item

Location:
Ratje

Notes:
Enter private notes here. These will only be visible by ELAR staff members and the depositor.
Interface-located functions

- see usage data/statistics
Interface-located functions

- apply for access at resource
User
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
Interface-located functions

- ‘emergency button’
Documentation of Reyesano

Found 245 bundles in this deposit (page 1 of 31)

1 2 3 4 5 6 7 8 9 ... next » last »

Abraza [8 August 2012, 6:02 pm]
Personal history
Recorded on 24 May 04
Personal story. Encounter with a jaguar during a fishing expedition

Edit bundle Hide bundle
Keywords: Reyesano - Personal history - fishing expedition - jaguar - Gumercindo Yanana

Adelanto [8 August 2012, 6:02 pm]
Interview
Recorded on 13 July 2005
Antoine interviews Teodilo who talks (negatively) about the life in Reyes, about how poor they are, how nothing happens. Teodilo also talks about the path to Rurrenabaque when there was no highway.

Edit bundle Hide bundle
Keywords: Reyesano - Interview - life in Reyes - path to Rurrenabaque - poverty - Antoine Guillaume - Teodilo Manoi

Alforjas [8 August 2012, 6:02 pm]
discussion
Recorded on 13 July 2005
Berna talks about traditional bags and clothing that her ancestors' grandmothers would weave.

Edit bundle Hide bundle
Found 245 bundles in this deposit (page 1 of 31)

1 2 3 4 5 6 7 8 9 ... next > last »

- **Abrazar [Hidden]**
  - Personal history
  - Recorded on 24 May 04
  - Personal story: Encounter with a jaguar during a fishing expedition

- **Adelanto [8 August 2012, 6:02 pm]**
  - Interview
  - Recorded on 13 July 2005
  - Antoine interviews Teófilo who talks (negatively) about the life in Reyes, about how poor they are, how nothing happens. Teófilo also talks about the path to Rurrenabaque when there was no highway.

- **Alforjas [8 August 2012, 6:02 pm]**
  - Discussion
  - Recorded on 13 July 2005
  - Berna talks about traditional bags and clothing that her ancestors’ grandmothers would weave
Towards portal services

- an example of a portal approach to widening participation – Participant-Driven Language Archiving ‘That’s me’ (E Garrett)
- a proposed portal for directly implementing community relationships to materials held in language archives. People can
  - claim roles
  - ‘collect and correct’
  - add information
Participant-Driven Language Archiving

- follows from Community-Based Language Research (Czaykowska-Higgins), ‘participatory archives’, ‘community curation’. Mary Linn’s proposed ‘Community-Based Language Archive model’ (Linn, forthcoming)

“the archive actively engages with the relevant community in conducting all levels of documentation, describing and contextualizing, maintenance of, and dissemination of information”
Implications of ‘That’s me’

- widening usage and potential new audiences through increased traffic from ‘That’s me’ portal
- enhance, extend, improve quality and contextualisation of materials
- if archives subscribe/collaborate, then new paths for managing access
- demonstration of our declarations about audiences, and help implement goal of serving them
Conclusion: towards portal services

- existing solutions don’t seem so friendly to scaled implementation of managed access, or to non-academics
Conclusion: towards portal services

- access portals:
  - need scaled managed access (taking into account many relevant factors)
  - implement *trusted relationships* between resources, communities, users, depositors, and archives
  - foster and utilise other-sourced associations between materials (e.g. ‘I also said this’)
  - could be from diverse parties serving various needs and values, in a wider ecology
References


Linn, M. forthcoming. Living Archives: A Community-Based Language Archive Model, in D Nathan & P Austin (eds) *Language Documentation and Description* vol 12


THANK YOU!

www.elar-archive.org

David Nathan   djn@soas.ac.uk
Kakia Chatsiou  kakia.chatsiou@soas.ac.uk