Multilayered Collaboration: Documentation & Archival Construction in the Aftermath of the 2015 Nepal Earthquakes

Kristine Hildebrandt, Tanner Burge-Beckley, Jacob Sebok
Southern Illinois University Edwardsville
Slides available at https://tinyurl.com/yxkkobj3
Archiving Documentation Materials: The Common Scenario

- The researcher and/or community activist records language data from a single language, representing different genres/contexts
- Once documentation and data preparation are completed, the materials are sent to an existing (external) repository, which “ingests” the materials for long-term storage and preservation
- Access can be controlled and restricted according to community concerns and considerations
- **Our project:** collect as many instances as possible of survivor narratives from speakers across myriad **Tibeto-Burman** ethnolinguistic communities representing multiple geographical regions of Nepal at a single point in time

Recording a Nubri narrative in Gorkha District
The Aftermath of the 2015 Earthquakes

- The first earthquake (4/25, M 7.8) killed 9,000, with an epicenter in Gorkha
- The second (5/12, M 7.3) killed an additional 200
- Our immediate concern (after learning about survivors, welfare of friends and relatives): capturing and preserving survivors' nuanced and diverse experiences of living through a natural disaster, in their own languages, before these narratives were generalized and altered for a global mass media audience
- We secured funding from the NSF RAPID mechanism (BCS 1547377: Narrating Disaster: Calibrating Causality & Responses to the 2015 Earthquakes in Nepal)

- We hoped survivors & responders would recount their reactions at the time of the events, and also describe what they know about earthquakes more generally, based on lifelong experiences & belief systems
The Linguistic Situation in Nepal

- Home to over 100 languages representing 5 families
- For many languages, the first major collection of transcribed, translated, and annotated materials with situational metadata and access to original audio and video.
- The materials center on theme of narration of survival & impact, in languages for which there is otherwise very little available description or data
- Exploration of a discourse genre that has barely been examined (but cf. Cox et al 2008 for media accounts of disaster narration; Baker et al 2008 & Hansen 2006 for refugee and foreign policy narratives) and allows further research on how survivor experiences map onto the physical world, and how this is linguistically encoded in terms of source, emotional experience, reactions to that experience, and also priorities identified by experiencers

<table>
<thead>
<tr>
<th>Language, Glottocode, ISO 639-3</th>
<th>Sub-grouping</th>
<th>District</th>
<th>EGIDS Scale (Lewis &amp; Simons 2010)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nubri, nubri1241, kte</td>
<td>T-B: Bodish, Gyalshundo-Nubri-Kyirong</td>
<td>Gorkha</td>
<td>6a Vigorous</td>
</tr>
<tr>
<td>Tsum, tsun1240, itz</td>
<td>T-B: South-Western Tibetic</td>
<td>Gorkha</td>
<td>6a Vigorous</td>
</tr>
<tr>
<td>Kutang, kuta1241, ght</td>
<td>T-B: Bodish, Kaise-Ghale-Tamangic</td>
<td>Gorkha</td>
<td>6a Vigorous</td>
</tr>
<tr>
<td>Gurung, west2414, gvr</td>
<td>T-B: Bodish, Tamangic</td>
<td>Manang</td>
<td>6b Threatened</td>
</tr>
<tr>
<td>Manange, mana1288, nmm</td>
<td>T-B: Bodish, Tamangic</td>
<td>Manang</td>
<td>6b Threatened</td>
</tr>
<tr>
<td>Nar, nar1239, npa</td>
<td>T-B: Bodish, Tamangic</td>
<td>Manang</td>
<td>6a Vigorous</td>
</tr>
<tr>
<td>Pha, nar1239, npa</td>
<td>T-B: Bodish, Tamangic</td>
<td>Manang</td>
<td>6a Vigorous</td>
</tr>
<tr>
<td>Lowa, lowa1242, loy</td>
<td>T-B: South-Western Tibetic</td>
<td>Mustang</td>
<td>6a Vigorous</td>
</tr>
<tr>
<td>Tibetan, tibet1272, bod</td>
<td>South-Western Tibetic, further</td>
<td>Mustang</td>
<td>6a Vigorous</td>
</tr>
<tr>
<td>Baragaon Tibetan, baral1356</td>
<td>Tibetic, closely related to Lowa</td>
<td>Mustang</td>
<td>Unknown</td>
</tr>
<tr>
<td>Dokdeh/*bregs skad</td>
<td>South-Western Tibetic, further</td>
<td>Mustang</td>
<td>Unknown</td>
</tr>
<tr>
<td>Tibetan, no known identifiers</td>
<td>classification uncertain</td>
<td>Mustang</td>
<td>Unknown</td>
</tr>
</tbody>
</table>

Languages in this Project
RAPID Project Organization

- We recruited 16 local residents who spoke the languages and who also had basic computer skills
- We trained them on A-V recording techniques, back-up, oral informed consent, and we co-constructed prompts to gather narratives (and in Mustang: more formal interviews)
- A day-long “crash course” in field methods in Kathmandu
- The residents returned to their homes over the coming weeks (as roads & footpaths became accessible); they also collected narratives from diaspora households in Kathmandu (also severely impacted)
Archival Workflow

- In Nepal, recordings transcribed, translated (Nepali, English), and a sub-set had morpheme interlinearization.
- We ended up with 96 fully or partially transcribed narratives (varying from 3-15+ minutes in length) across 8 languages.
- Participants ranged from local residents (often agricultural or local civil servants) to members of the Buddhist religious community, to returning “diaspora responders.”
- Data transfer to home institution in subsequent trips to Nepal.
- UVA A-V and transcript proof-reading, cleaning-up, and materials upload done by PI and SIUE volunteer RA’s.
University of Virginia’s Tibetan & Himalayan Library

- Hosts video narratives and transcripts into six collections: Nubri, Tsum, Kutang, Lowa, Nar-Phu, Manange
- Developed with a CMS designed by THL librarians
- Platform based on Drupal, an open source software for websites & other digital applications
- The metadata encoding worked with both universal standards, and standards specific to the “THL-verse”
- Each language is a unique collection; each collection houses narratives plus transcripts.
- Within each collection, items catalogued with item and content metadata, as allowed by oral consent.
- By clicking on the item: A thumbnail view of the video & time-aligned transcript. The video can be played in its entirety, or else individual transcript lines can be clicked on to hear that utterance in isolation
The Need For Another Repository, And an Idea....

- This project was assembled quickly (3 months from RAPID application to equipment purchase, travel to Nepal, recruitment & training of resident field assistants)
- We ended up with a variety of other materials that needed cataloguing & storage, but these were not in line with what we had originally agreed to deposit in THL: many still images, over 100 formal Mustang interviews, derivatives
- SIUE Interdisciplinary Research & Informatics Scholarship (IRIS) Center
- We have an ITS issuance of Omeka, and guarantee of regular backups
- But: Omeka is an “out of the box” application, requiring on-site construction
- We also wanted an online web exhibit interface that would link the two collections (UVA THL & our Omeka collection) along with other information about Nepal (language situation, post-earthquake updates)
What Now?

1. Requirements for Archive and Exhibit
3. Choose Platforms to Facilitate Requirements
4. Construct Site Architecture (Shells)
5. Populate the Site with Data
Choosing the Platform(s)?

- Aesthetic Appeal
- Highly Customizable
- User-Friendly Interface
- Graphics-Driven Experience
- Respected Platform
- Dublin Core Metadata Standard
- Built to House Archives
- Infrastructure Already in Place
What is Omeka?

- Free
- Open Source
- Designed for Exhibits
- Dublin Core Metadata Standard
Information Architecture

We needed to organize

- Images,
- Audio Files
- Transcripts

According to

- Region
- Language
CSV Import - A Miracle

1) Import data to spreadsheet
   a) Mass-manipulation of data
   b) Sorting / Filtering
   c) Error Checking

2) Export spreadsheet as .CSV
   a) Single Location

3) Upload .CSV to Omeka
   a) Minimizing Potential User Error
The Archive

Each language had both **Audio Files** and **Transcripts**.

In order to facilitate a productive user experience, we adopted a **Trinomial** naming schema that identified:

- Interview Language
- Reference ID
- Media Type
The Finished Product

With Omeka, a great deal of Metadata is visible to every user with the click of a button.

But how to make this accessible to a wider audience?
Archive vs. Exhibit

- Not all of the research data is approachable to a wider audience
- A exhibit provides an accessible portal of contextual information for those unfamiliar with the research.
Wordpress and Omeka: Two Sides to a Coin

- For general information presentation, Wordpress is an open source, user friendly choice to build a complementary exhibit.
- Like Omeka, is built on PHP as a server side script development language.
- Provides a different form of CMS than Omeka.
Cohesive Presentation

● Both Omeka and Wordpress provide free themes to customize color and arrangement presentation of page information
● With two complementary installations, we chose themes that were built on the same core Bootstrap framework
● Themes were then customized directly to appear more similar
PHP Customization

- Theme customization involves PHP source editing
- Color and style customizations were made in CSS modification
- CSS and PHP knowledge can be employed from a knowledgeable Computer Science student or professional web developer
Benefits of a Self Hosted Archive

- Uses university resources
- Customization of archive design
- Set and uphold metadata standards
- Opportunity for cross-department collaboration

https://iris.siue.edu/nepal-earthquakes-archive/
We Also Faced Challenges

- We had to construct informed consent protocols to not compromise the dignity of participants in overwhelmingly complex and difficult situations
- We had to be sensitive to local administrative policies
- Our field researchers worked with a variety of equipment types, and often had to share resources
- They also struggled with aligned file naming conventions, and many transcript files needed to be transcoded
- Some of this stemmed from different disciplines coming together, and some of it came from the different field research team choices, issues of timing, issues of material access, and the constant tectonic, population, and political shifting that followed in the weeks and months after the earthquakes
- We also had challenges in terms of pairing mis-matched transcripts, with audio and video files, or A-V with no transcription whatsoever
Final Thoughts

- This presentation does not imply that documentation initiatives should attempt home-designed and locally-housed archival repositories instead of other, well established archives.
- Established archival repositories have an advantage of a knowledgeable steering committee, trained staff, established standards for data formatting, encryption, preservation/protection, and of course a great deal of storage space.
- We had residual materials and the chance to make use of local resources as a teaching and learning opportunity, and we have home institutional support.
- We turned to established archives for inspiration and guidance.
- We hope that by documenting our challenges and our solutions in the construction and customization of these two digital tools, other users will be able to adopt and adapt in their own endeavors.

https://tinyurl.com/yxkkobj3
Thank you and Questions

- NSF BCS 1547377 funding, including two REU supplements
- We are grateful for the participation of: Geoff Childs, Sienna Craig, Mark Donohue (co-I’s); Dr. Dubi Nanda Dhakal, Mr. Bhoj Raj Gautam (managers); Lhakpa Hritar, Lhakpa Norbu, Jangchuk Sangmo, Nyima Samdrub, Tenzin Kunchok, Namgyal Ngudup, Pema Norbu, Tenzin Gyaltsen, Ngawang Tsering Gurung, Yangjin Bista, Karma Choedon Gurung, Tsewang Gyurme Gurung, Sophiya Lama, Chimi Lama, Tsering Topden Gurung, Anjana Ghale (field researchers); Sarah Song, Kristin Kaskeski, Ishu Jha, and Pratik Lamsal (volunteer SIUE R.A.’s who assisted with UVA data transfer)

https://tinyurl.com/yxkkobj3

Video training in Kathmandu (July 2015)

Co-I S. Craig at project team check-in (October 2015)