



# Language Archiving

## Where we've been and where we're going

Gary Holton *University of Hawai'i at Mānoa*

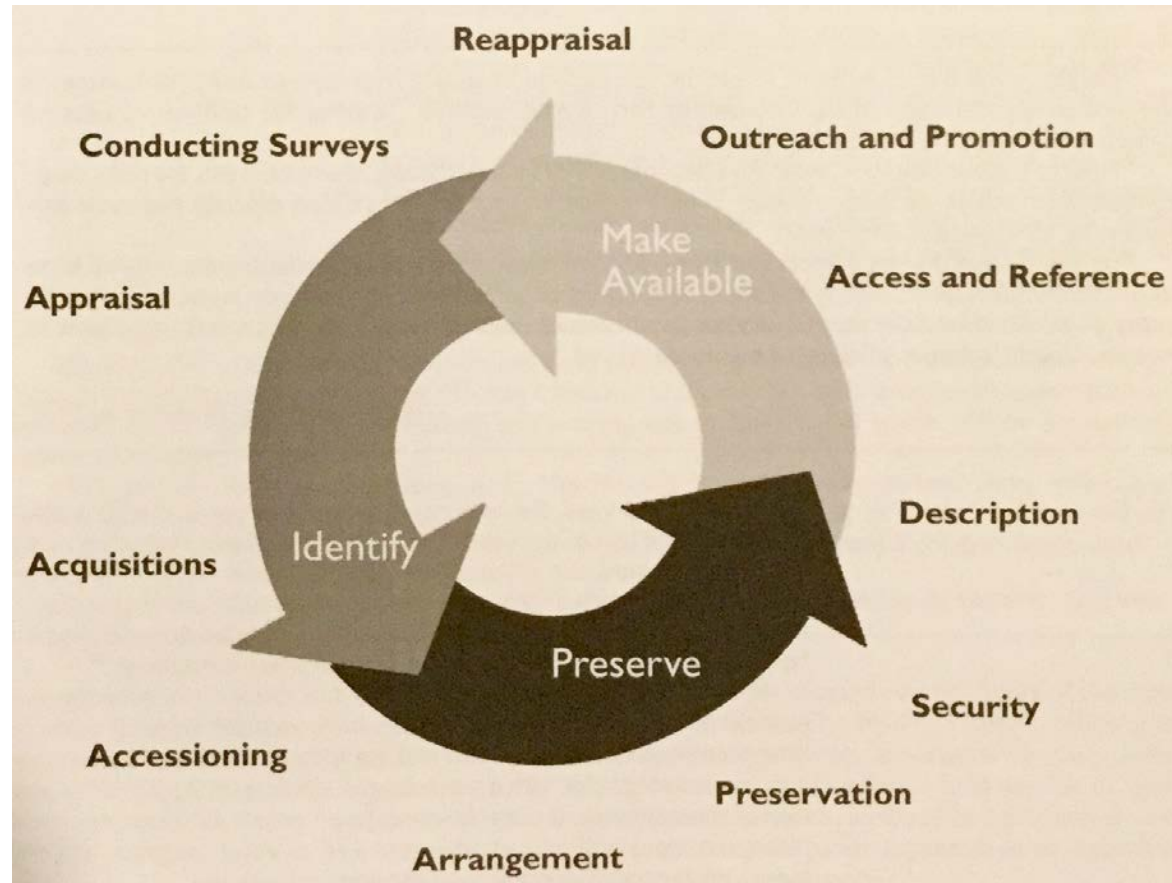
Workshop on User-Centered Design of Language Archives, Feb20-21, 2016



This material is based upon work supported by the National Science Foundation under Grants No. BCS-1543763 and BCS-1543828.

# What is an archive?

- Materials
- Location
- Institution
- Archival mission
  - Identify
  - Preserve
  - Make available



(Hunter 2003)

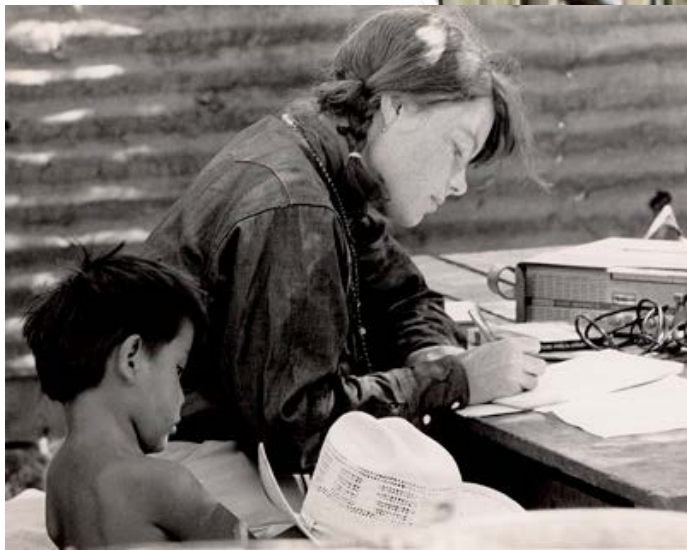
# Brick & mortar language archives

1743 American Philosophical Society

1879 National Anthropological Archives

1953 Survey of California Indian Languages

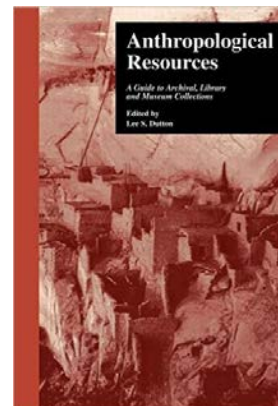
1972 Alaska Native Language Center



# American Philosophical Society



- “The first drudgery of settling new colonies is now pretty well over, and there are many in every province in circumstances that set them at ease, and afford leisure to cultivate the finer arts, and **improve the common stock of knowledge.**” (Franklin 1743)
- Led to creation of the American Philosophical Society, which eventually came to house some of the most important collections relevant to American languages
- Franz Boas Papers and Field Notes (73 lin. ft.)



# Melville Jacobs papers

## UW Libraries Special Collections

- 1902-1971, 78.23 cubic feet
- Many languages today known only through Jacobs' work



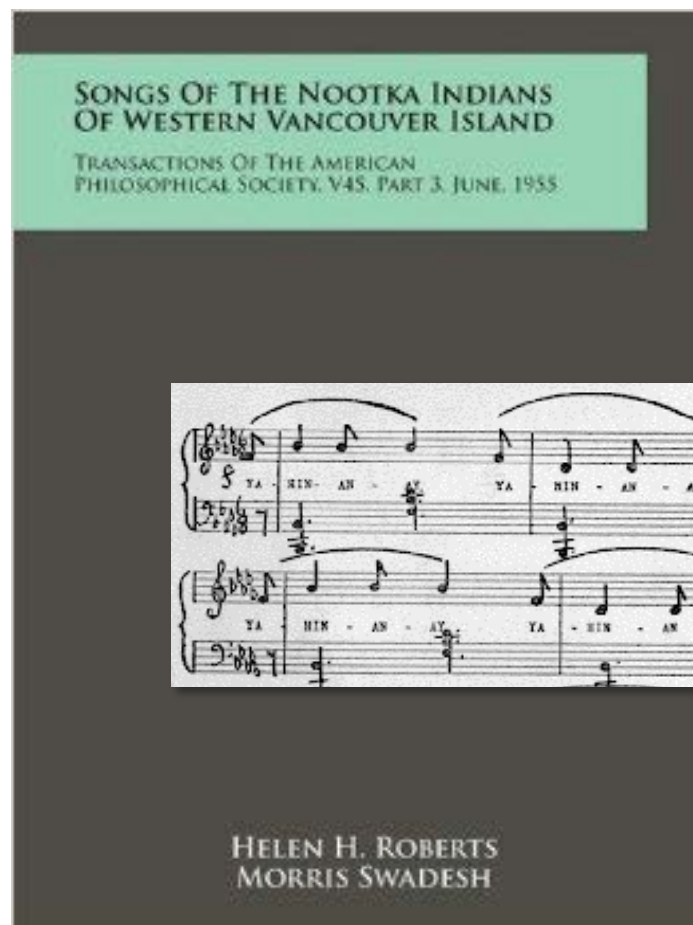
Property of Special Collections, University of Washington Libraries.



# An expectation of archiving

- Few references to archiving during the “golden age” of Americanist linguistics
- Linguists were actively creating a documentary record; the assumption was that materials would be preserved
- The legislation which founded the ANLC to “study the languages native to Alaska” made no specific mention of archiving. It didn’t need to. For what would be the purpose of collecting documentation of endangered Alaskan languages if there were no intention of preserving the documentation itself?
- Language archiving closely tied to documentation

# Access / Utilization



- Must be viewed in historical context
- Nootka songs, recorded by Sapir 1910, 1913-1914
- Reviewed by Helen Roberts with speaker Alex Thomas in 1934
- Published 1955, Roberts & Swadesh

“the culture which [Sapir] describes in his notes of 1910 and 1913-14 is today only a vague memory” (Gunther 1957)

# Emergence of digital language archiving

## Response to three independent developments

1. Renewed attention to endangered language documentation following 1991 LSA
2. Digital data transformation
3. Urgent need to access language resource for language maintenance





# A renewed focus on documentation

- The documentary focus of linguistics in the first half of the 20th century was largely forgotten by mainstream linguistics in the 1960's as the profession became fascinated by theoretical models
- “Obviously we must do some serious rethinking of our priorities, lest linguistics go down in history as the only science that presided obliviously over the disappearance of 90% of the very field to which it is dedicated” (Krauss 1992:10)
- Two decades later the Krauss' clarion call has largely been heeded

# Digital transformation



- As linguists rushed to record endangered languages, technology was evolving so quickly that recording devices specified in a grant proposal became obsolete before the proposal was funded
- Linguists were thus forced to grapple with technological standards in order to ensure that the records they were creating would be of lasting value
- An exponential increase in the ability to create language documentation data in digital form has required the development of standards for formats and metadata so that digital data can be effectively managed, thus avoiding “digital carnage” in which data become effectively inaccessible (Bird & Simons 2003)
- Traditional archives simply weren’t ready for digital data.

# Integration of archiving with documentation

- Real-time archiving
  - “Archiving directly from the field” (Robinson 2012)



Contra archivists' claim that "Archival materials are never explicitly created." (Hunter 2003:8)

# Support for language conservation

- “it is important to add that California's Native people and languages are resilient and adaptive; that languages are being learned, transformed, and revived in the face of tremendous obstacles; and that so far, **every generation of scholars has been wrong in its prediction that California languages are on the verge of extinction.**”

I HEAR THEIR VOICES.  
THEIR FACES I CANT  
SEE. THEY HAVE MUCH  
TO TALK ABOUT. AND  
THEY WANT TO  
SPEAK WITH ME.



# Support for language conservation

- Language conservation efforts require immediate access to documentary material
- In theory digital preservation depends crucially on access, namely, “the storage, maintenance, and *accessibility* of a digital object over the long term” (NINCH 2003:199)
- In practice the two goals may be at odds, as the desire to create innovative and attractive products such as multimedia websites and mobile apps favors reliance on proprietary, non-archival formats.



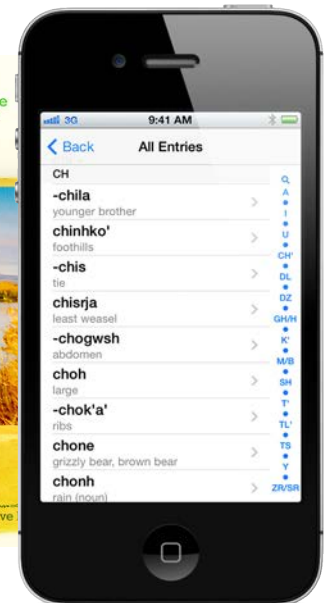
# Balancing preservation and access

- Preservation

- Cataloging
- Re-foldering, boxing
- Digitization
- Long-term storage
- Archival formats

- Access

- Linked media
- Web technologies
- Mobile apps
- Convenient formats





# Linguists go it on their own:

## The emergence of digital language archives

- 2000 Archive of the Indigenous Languages of Latin American (AILLA)
- 2000 Dokumentation bedrohte Sprachen (DoBeS)  
[rebranded The Language Archive (TLA) in 2011]
- 2003 Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC)
- 2004 Endangered Languages Archive (ELAR)
- 2008 Kaipuleohone Digital Language Archive

# Development of digital standards and best practices for language archiving

- 2000, Workshop on Web-based Language Documentation and Description, U Penn
- 2001, Open Language Archives Community (OLAC)
- 2002, Electronic Meta-Structures for Endangered Languages Data (E-MELD)
- 2003, Digital Endangered Languages and Musics Archiving Network (DELAMAN)
- 2007, ISO 639-3 language codes

*All of these efforts led by linguists working outside the field of archives*

# Linguists turned the field of archiving on its head

- Focus on resources related to a particular language rather than on the intellectual legacy of a particular individual (contra *respect des fonds*)
- Documentation for the purposes of archiving (creating an archival record)
- Focus on “real-time” archiving

# Meanwhile traditional archives have evolved

- Digital best practices, such as National Digital Stewardship Alliance
- ISO 16363 Trusted Digital Repositories
- Greater reliance on content specialists
- Move toward participatory archiving, involving depositors and users as collaborators

# Increased focus on usability

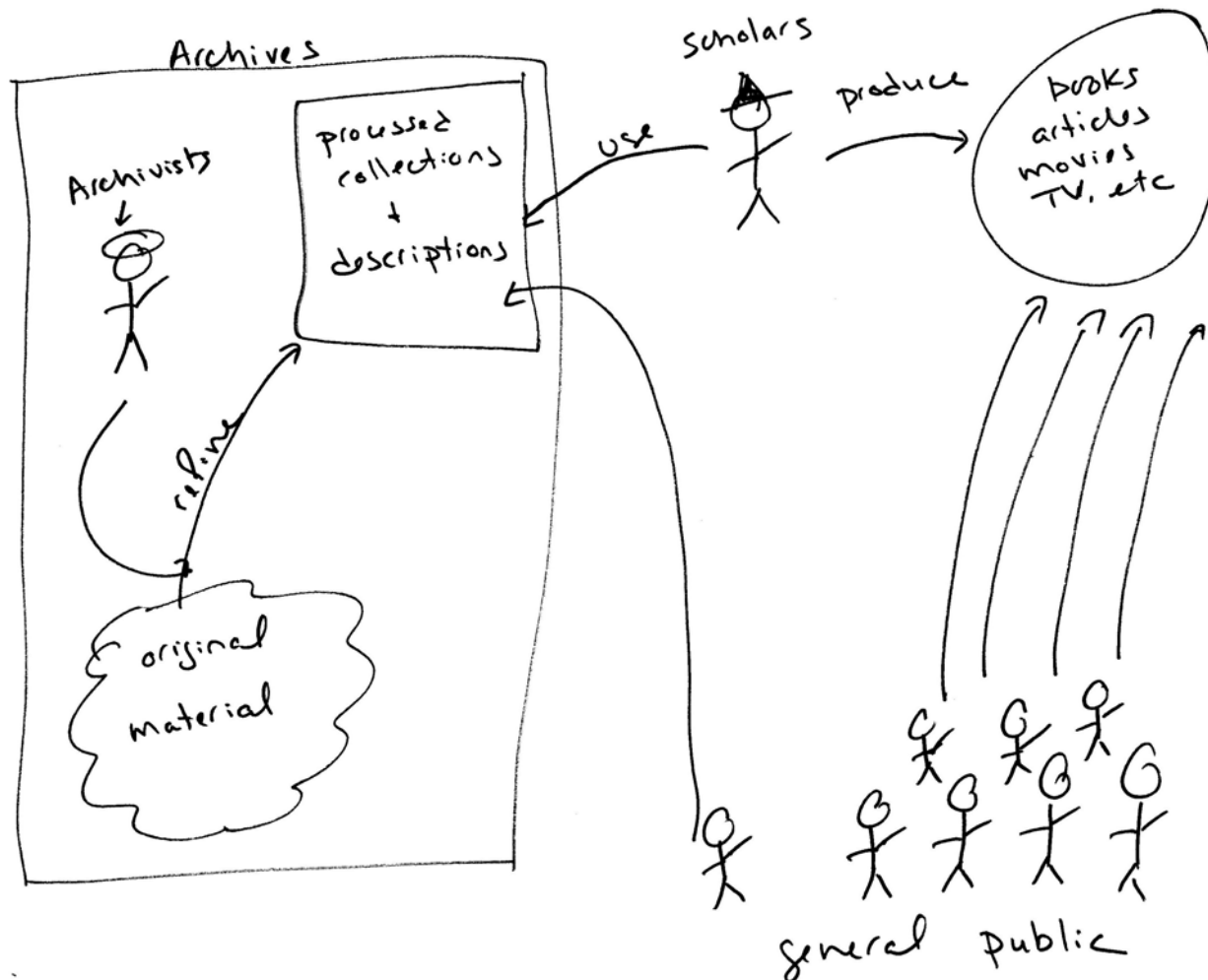
- “In the traditional archival paradigm, the first issue has been preservation and the archival process. Usability of the archive has generally been of lesser importance until recently.” (Huvila 2008)
- In a **participatory archive** “usability does not denote use alone, but also denotes a deeper level of involvement in the sense of actual participation in the archive and in the archival process.” (Huvila 2008)

# Increased focus on the user

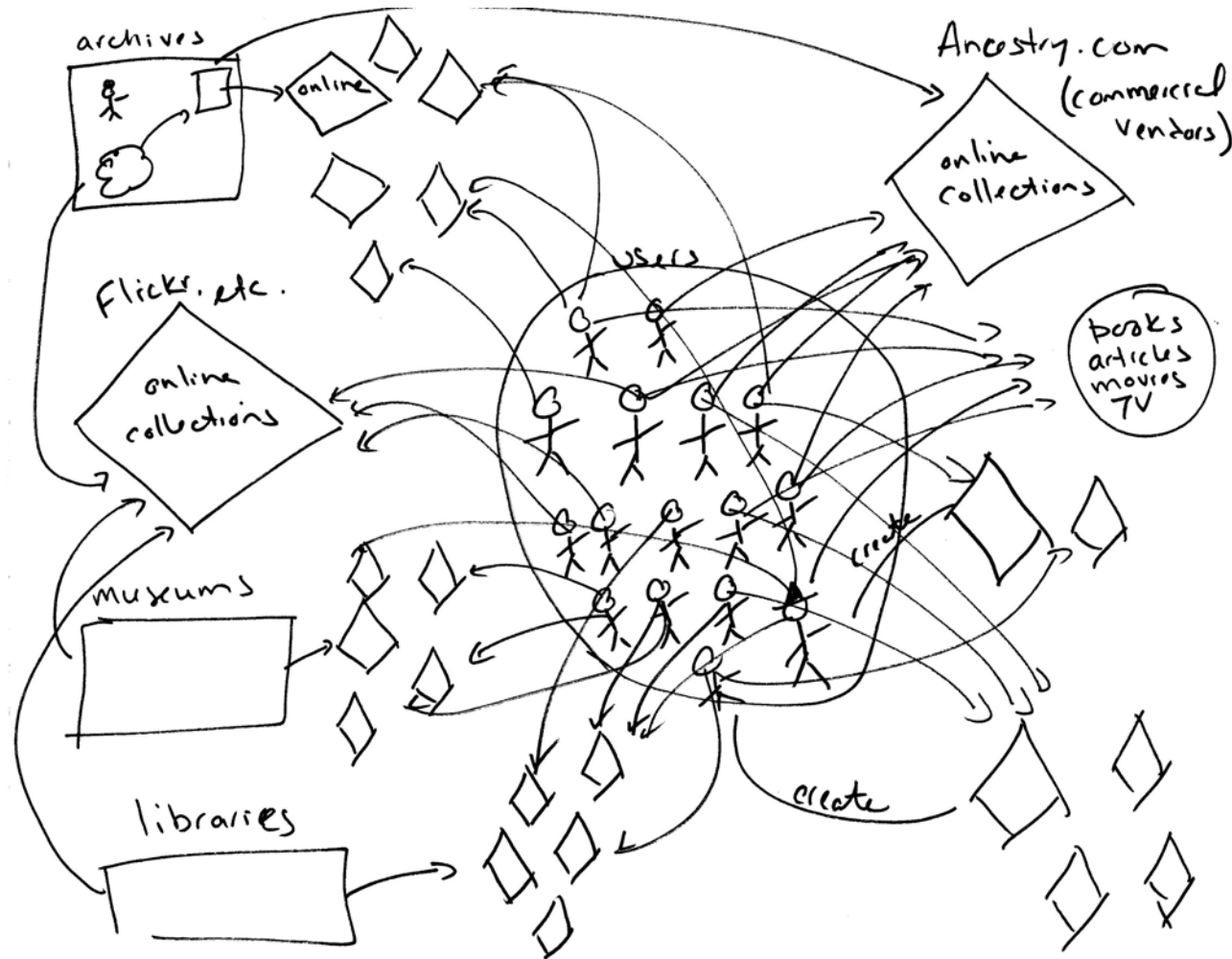
- “Newer forms of electronic archiving restore the deep link of the archive to popular memory and its practices, returning to the non-official actor the capability to choose the way in which traces and documents shall be formed into archives,” (Appadurai 2000)
- “It is important to recognize that [online users] are still users of our services and that they need to be taken into consideration in the same way as the more tangible users who occupy the tables in our searchrooms.” (Hill 2004)
- “Re-envisioning archival principles of appraisal, arrangement, and description to actively incorporate participation from traditionally marginalized communities will ... allow these communities to preserve empowered narratives.” (Shilton & Srinivasan 2007)



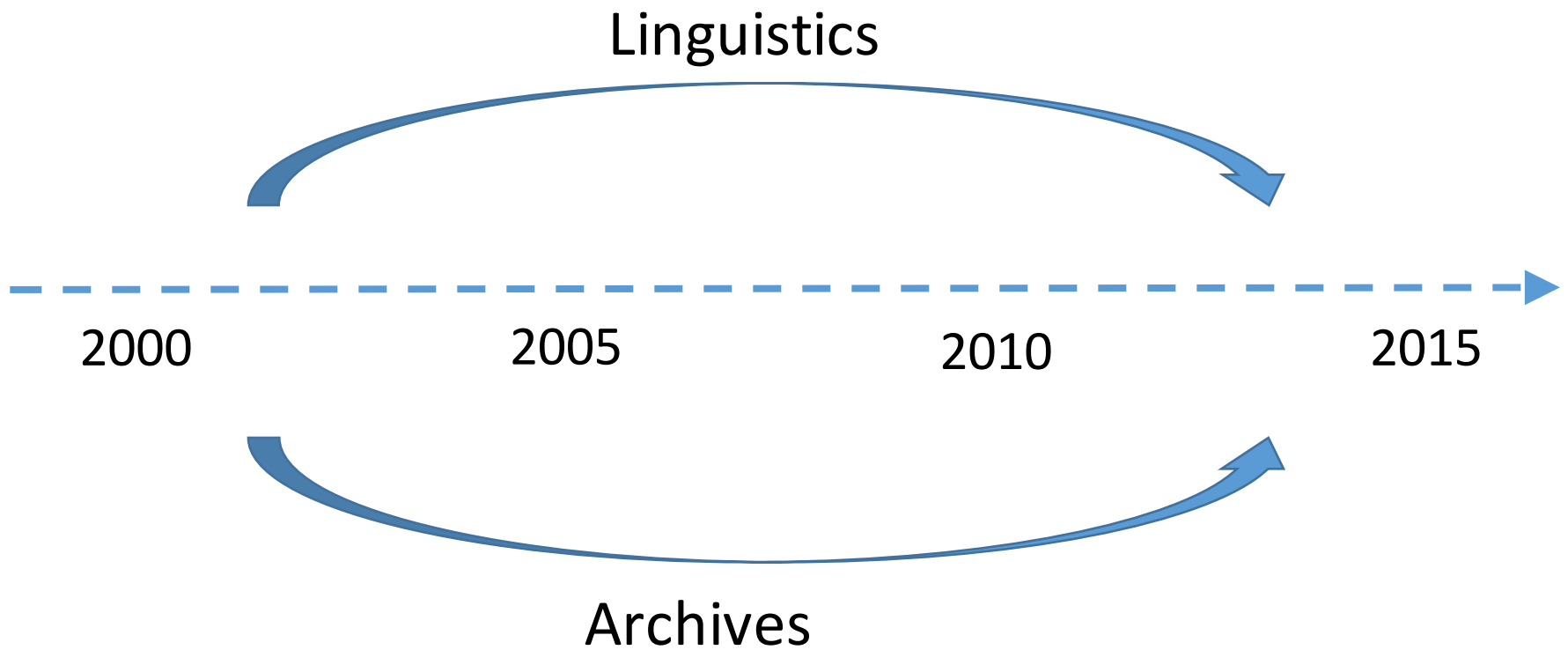
# From traditional mediated archives



# Toward participatory archives

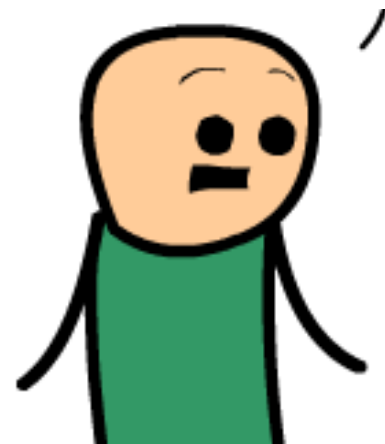


# Parallel evolution



# Whither digital language archiving?

- Brick & mortar archives are catching up, solving many of the preservation and access problems that digital language archives set out to solve
- Has digital language archiving thus become obsolete, to be replaced by mere “archiving,” where digital preservation is the norm and language resources are treated just like any other resource? (cf. Holton 2014)
- Are language resources “special”?



# Access and digital language archives

- Access is a fundamental part of archiving
- Now in the post-digital revolution era, there is a renewed push for access
  - It's easy: digital return, digital repatriation
  - Supports a return to data-driven linguistics (reproducible research), drawing on larger data sets
  - Support for language conservation

# How are digital language archives used?

- Vast majority of users are heritage language learners (Austin 2011)
- Less interest in traditional language documentation
- More interest in cultural content (cf. Holton 2012)
  - Songs, stories
  - Learning materials (“phrasebooks”)
  - Biographical/genealogical information
  - History
  - Traditional knowledge
- Often struggle to interpret archival documents and recordings







# Materials may be inaccessible to the untrained user

[illegible]

on before the rest of  
with normal intervening

g <sup>w</sup> ola-	g <sup>w</sup> ula-	gola-
g <sup>w</sup> ola	g <sup>w</sup> ula	gola
g <sup>w</sup> ola	g <sup>w</sup> ula	gola
g <sup>w</sup> oli	g <sup>w</sup> uli	goli
g <sup>w</sup> oli-	g <sup>w</sup> uli-	goli-
g <sup>w</sup> u- gu-		

# Recordings may be inaccessible without interpretation

- Silas Solomon 1992 
- Coray 1954 



# Access is also about control

- “To be an Indian is to have non-Indians control your documents from which other non-Indians write their versions of your history.” (Hagan 1978)
- “The whole archive is an expression of our sovereignty.” (Alexander quoted in Shepard 2014)

# Participatory archiving means more than mere access

- “Our digital archives are only useful if they enable us to have analog meetings and connections between actual people conversing and sharing.”  
(Alexander quoted in Shepard 2014)





# Participatory archiving means more than mere access

Elders stress role of the archive in language continuity

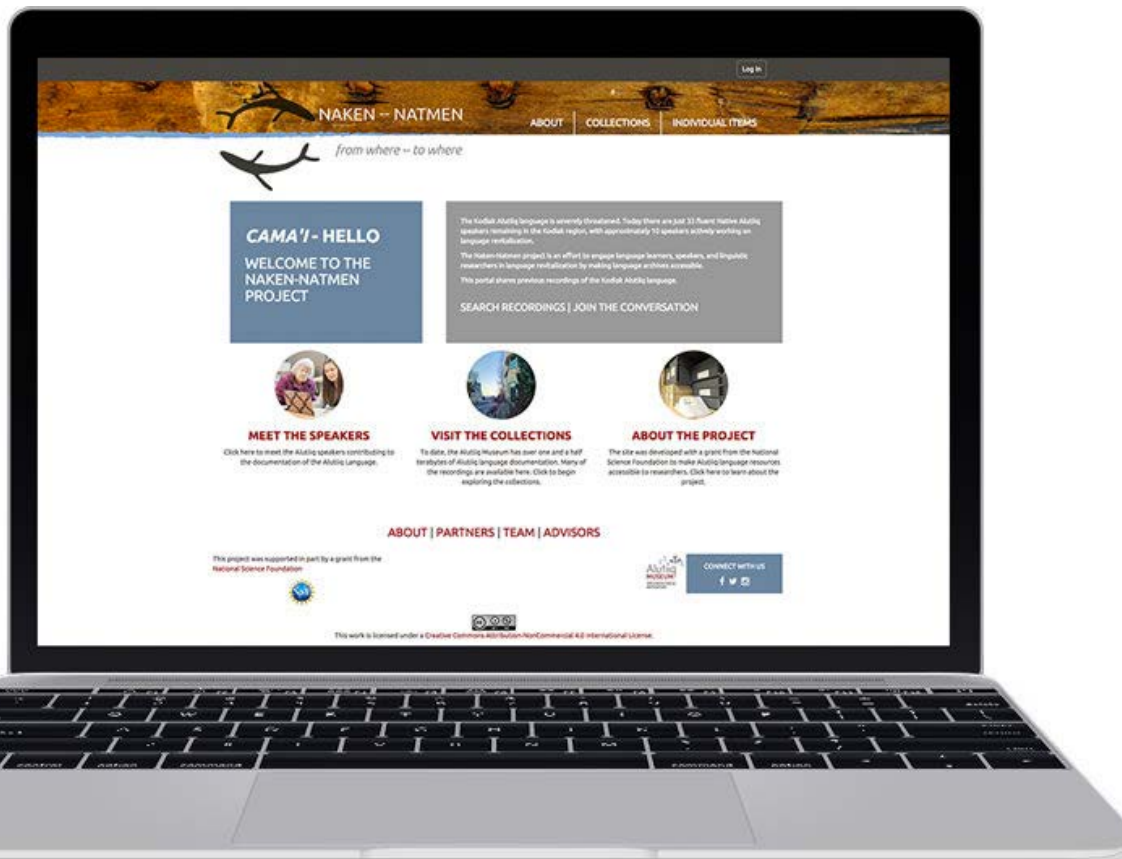


“If we don’t get it out and learn about it, where are we going to learn from? These are old recordings. We want to get it out and teach our younger children what the elder people are talking about. I think that’s a very good idea for getting it free so we can listen to them.”

*Andrew Balluta (2004)*

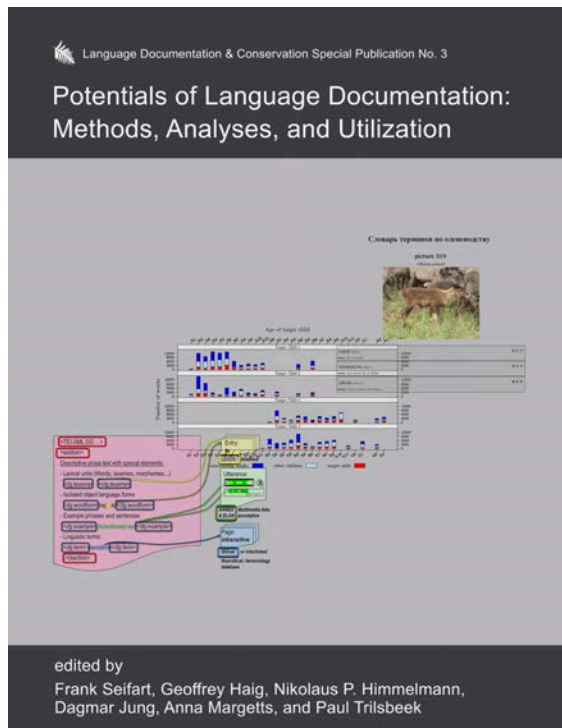
# From access toward accessibility

- Rise to alternate methods of sharing content, such as Mukurtu CMS





# How are digital language archives responding?



# Committee on Endangered Languages and their Preservation

# Utilization of Language Archives in Endangered Language Research, Revitalization, and Documentation (Portland, 2015)

# Potentials of Language Documentation (Leipzig, 2011)

# Community-based language archives

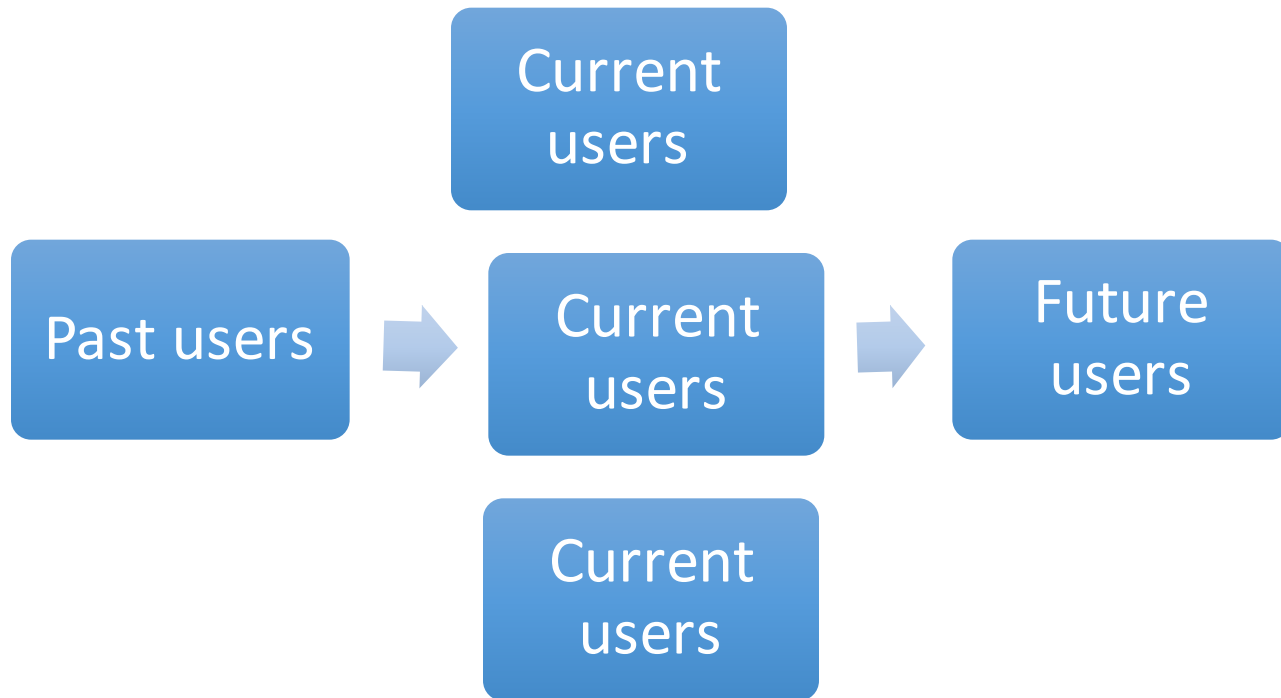
- “Archives have continuously evolved to remain relevant even as media environments, access concepts, and user demographics have changed. Endangered language archives are well-placed to participate in and articulate these shifts.” (Linn 2014)

# Problematizing “archives”

- The notion of “archives” – with a predefined mission independent of its user community – may itself be an example of top-down thinking.
- Rather than engaging user communities and asking them what they would like to see done with their language resources, we assume archiving – with its focus on preservation – is a universal good.
- User-centered design offers us an avenue to revisit these notions and engaging with user communities to develop better language archives.

# Who are the users?

- In order to apply User-Centered Design principles, we also need to know who the users are



- Austin, Peter. 2011. Who uses digital language archives? In *Endangered Languages and Cultures*.
- Bird, Steven and Gary Simons. 2003. Seven dimensions of portability for language documentation and description. *Language* 79(3).557-82.
- Hagan, William T. 1978. Archival Captive – The American Indian. *American Archivist* 41(2).135-46.
- Hill, Amanda. 2004. Serving the Invisible User: Meeting the needs of online users. *Journal of the Society of Archivists* 25(2).139-48.
- Holton, Gary. 2012. Language archives: They're not just for linguists any more. *Potentials of Language Documentation: Methods, Analyses, and Utilization*, ed. by F. Seifart, G. Haig, N.P. Himmelmann, D. Jung, A. Margetts and P. Trilsbeek, 111-17. (LD&C Special Publication no. 3). Honolulu: University of Hawai'i Press.
- Hunter, Gregory. 2003. *Developing and Maintaining Practical Archives*. Neal-Schuman.
- Huvila, Isto. 2008. Participatory archive: towards decentralised curation, radical user orientation, and broader contextualisation of records management. *Archival Science* 8.15-36.
- Linn, Mary. 2014. Living archives: A community-based language archive model. *Language Documentation and Description* 12.53-67.
- Robinson, Laura C. *Archiving directly from the field*.
- Shepard, Michael. 2014. "The Substance of Self-Determination:" *Language, Culture, Archives and Sovereignty*. University of British Columbia Ph.D. dissertation.
- Shilton, Katie and Ramesh Srinivasan. 2007. Counterpoint: Participatory Appraisal and Arrangement for Multicultural Archival Collections. *Archivaria* 63.87-101.
- Theimer, Kate. 2014. The Future of Archives is Participatory: Archives as Platform, or A New Mission for Archives. *ArchivesNext*.