

# Archiving and the work flow of field work

Nicholas Thieberger

# PARADISEC

Pacific and Regional Archive for Digital Sources in Endangered Cultures

# Language Archives

An integral part of language documentation

The locus for supporting documentary linguistic activity Ark - Hive (David Nathan, ELAR)

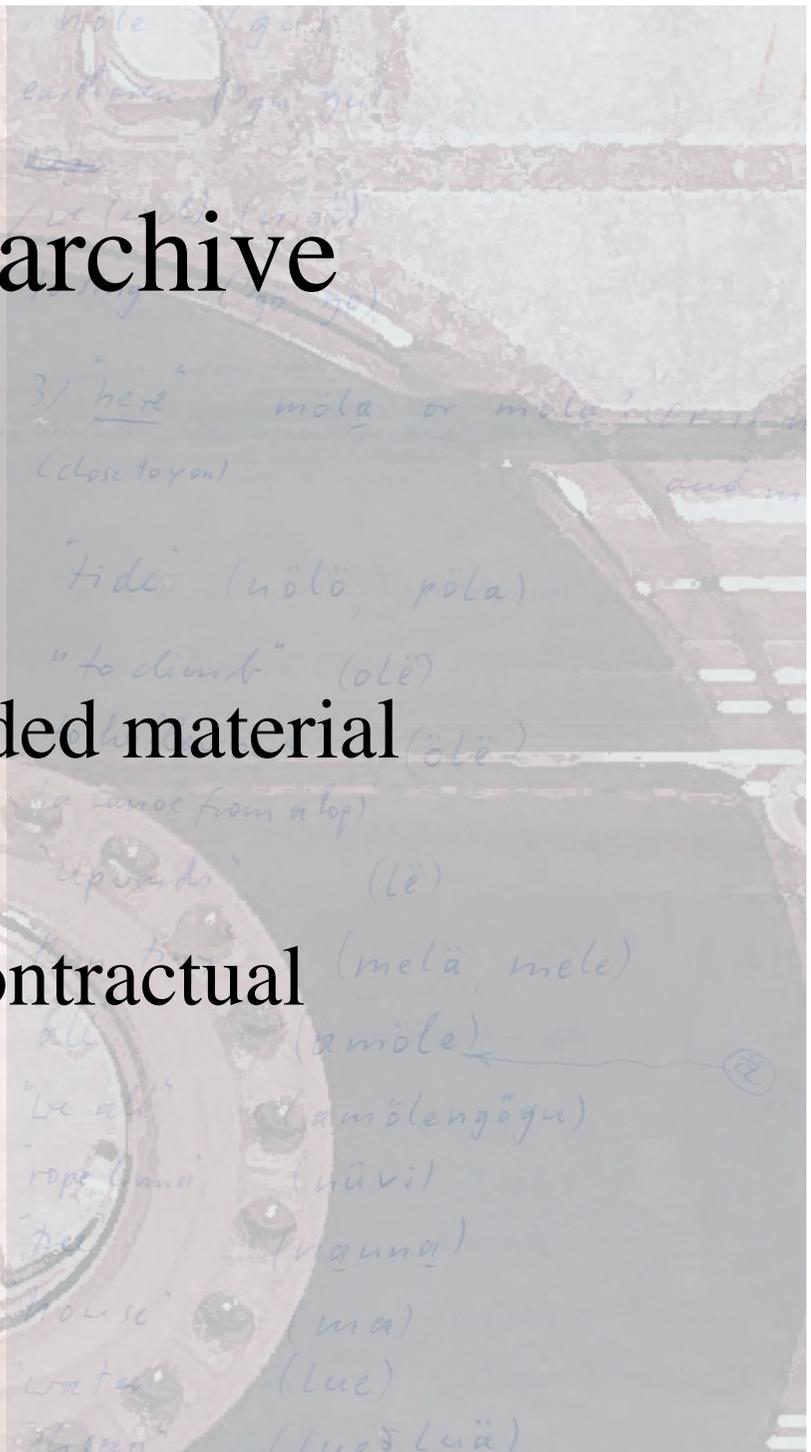
Need to develop archival methods for linguistic fieldworkers

# The coercive archive

Attached to a funding body

Obligatory deposit of recorded material

Enforces data formats by contractual requirement





[RNLD](#)  
[Rationale](#)  
[Goals](#)  
[Activities](#)  
[Mailing List](#)  
[FAQ](#)  
[Flier](#)  
[Convenors](#)  
[Advisory Panel](#)

email:  
[ReNeLD@hotmail.com](mailto:ReNeLD@hotmail.com)

## **RNLD: Resource Network for Linguistic Diversity**

*Networking practitioners working to record, retrieve & reintroduce endangered languages.*

This site will be developed to share skills on methods for working with endangered languages between language maintenance practitioners in the region broadly bounded by Australia, Indonesia, East Timor and Melanesia, including:

- documentation
- recording
- data management and archiving
- transcribing
- digitising
- training

Practical advice and links will also address techniques for retrieving linguistic information from archival sources and for programs to revive and reintroduce languages.

<http://www.linguistics.unimelb.edu.au/rnld.html>

# File naming conventions

Collection-item-file.extension

AB1-002-A.wav

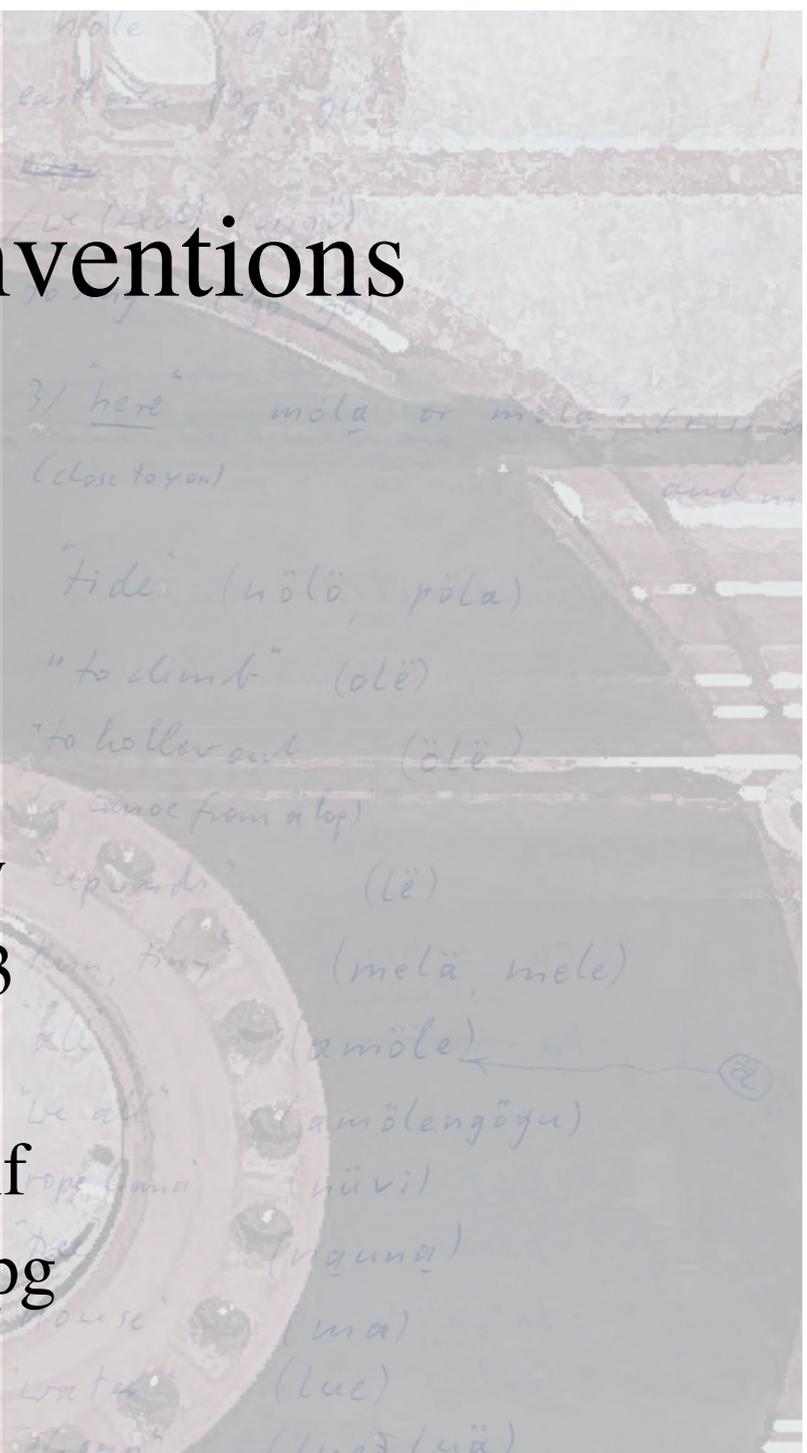
AB1/AB1-002/AB1-002-A.wav

AB1/AB1-002/AB1-002-A.mp3

AB1/AB1-002/AB1-002-A.xml

AB1/AB1-002/AB1-002-FN1.tif

AB1/AB1-002/AB1-002-FN1.jpg



(steer the canoe (ötätöä nöe lö nöguu)  
 (What is nöe in this? What is  
 tö in ötätöä exactly?  
 Also, what is tö in:  
 ayötöä nēwēngō | begin to work  
 apukötöä nao nöä | close the door  
 dwatöä bö ma | enter the house  
 maletöä müde | touch his arm  
 alvötöä natüm | learn your language  
 Also: what is lvē in alvötöä natüm - I learn your language  
 alvēä natölo lö yagüä | look after the sick man  
~~äpülvää nyö | blow the fire~~  
 daughter (inye or iayo - which is correct?)  
 mäle - brother - Is this brother of man, and  
 brother of a woman, or are these  
 different words for the brother of  
 a man, and that of a woman?

Page image of Stephen Wurm's fieldnotes on Aiwo, Solomon Islands, SAW2-018-00005

<hasTranscript>/<isTranscriptOf> proposed  
 addition to OLAC/Dublin Core <relation> element

# Everyday archiving

Creation of optimal archival forms of data through normal linguistic practice, e.g.

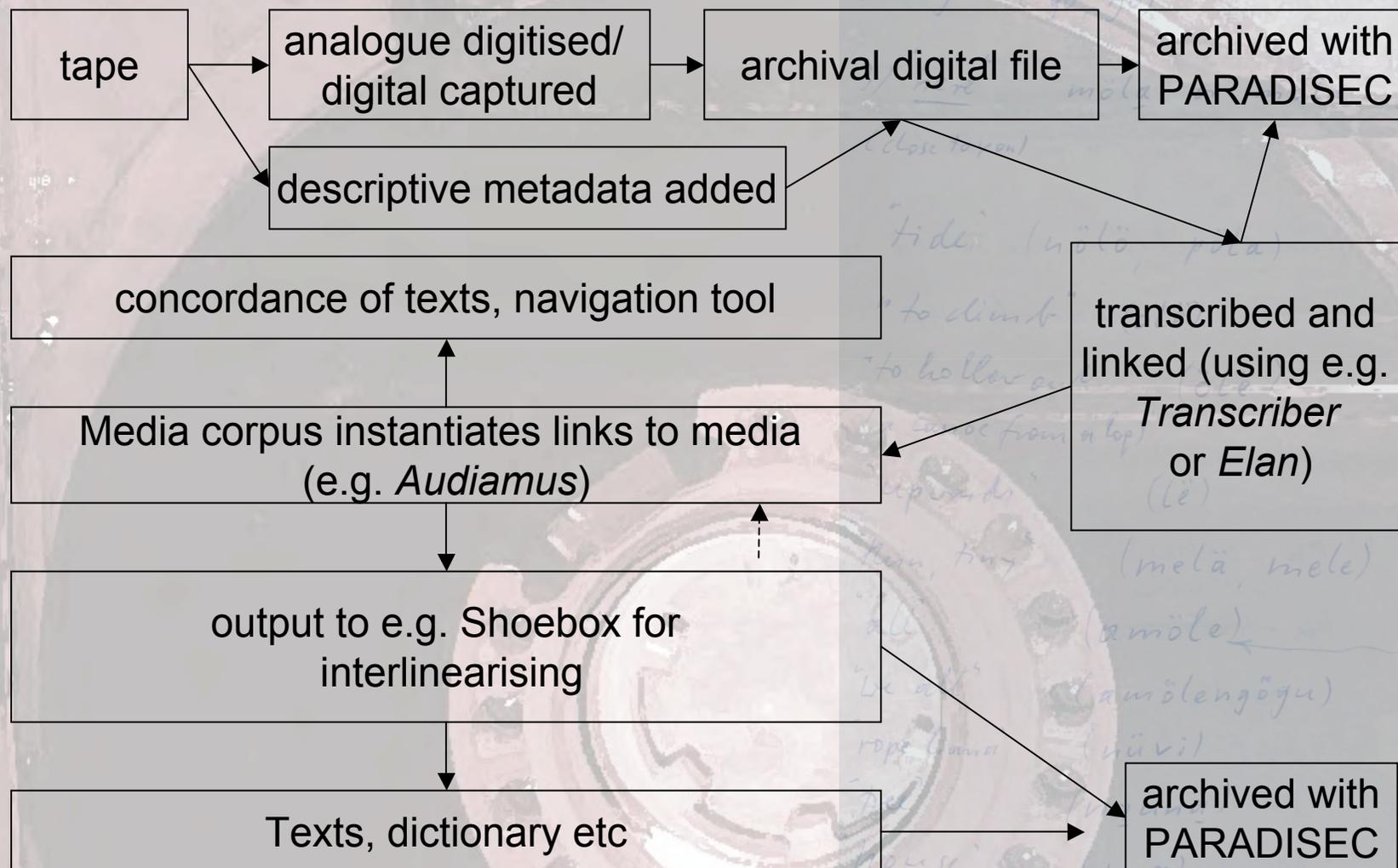
- Well-structured data (e.g. backslash codes, XML)
- Annotation with time-alignment
- Citation of archival data, which implies
  - persistent identification and location
  - interactive use of archival data

# Everyday archiving

Creation of optimal archival forms of data through normal linguistic practice, e.g.

- Tracking relationships within the data
  - speaker - tape - transcript - text
- Consent/deposit forms clarify intellectual property issues

## Example Workflow



# Contrast and compare

	Previously	Current
Data	Analog	Digital
Copyright in material clarified	Rarely	Consent forms signed by interlocutors (because deposit in an archive is envisaged as part of the process)
Filenames	Arbitrary	Persistent Identifiers

# Contrast and compare

	Previously	Current
Data structure	No explicit structure (implicitly marked by fonts and styles)	Explicit structure is used as the basis for derived forms (e.g. as in lexical files in Shoebox)
Archival accession of primary data	After use of the material by the researcher. (Typically post retirement or after death of the researcher.)	Incremental accession, ideally before use of the material by the researcher.

# Contrast and compare

	Previously	Current
Annotation of primary media	Little done, usually by hand.	More comprehensive annotation, using time-alignment and interlinearising.
Archival accession of annotations	Typically post retirement or after death of the researcher.	Work in progress archivable and overwritten by subsequent versions (safe backup)

# Contrast and compare

	Previously	Current
Persistent identification to support citation forms of data.	Maybe in fieldworker's notes, hampered by lack of discoverability.	Assigned by archive and persistent identifier resolved to an item in the archive.
Metadata standard	Library/MARC (large existing infrastructure)	DC / OLAC (support for small, collector-based archives)

# Contrast and compare

	Previously	Current
Metadata discovery	Library catalogs (not always interoperable)	Open Archives Initiative, subject specialised searches
Persistent identification to support citation forms of data.	Maybe in fieldworker's notes, hampered by lack of discoverability.	Assigned by archive and persistent identifier resolved to an item in the archive.

# Contrast and compare

	Previously	Current
Persistence of data	Analog tape in one location	Digital simulacra/copies (LOCKSS)
Relation between items	Ignored or treated in catalog	Treated in metadata and instantiated where possible (e.g. tape/transcript)

# Contrast and compare

	Previously	Current
Repatriation of copies	Copies of tapes provided from a single location.	Digital copies of tape/transcript in linked form. Available for download from the web.

# Conclusion

- Secure longterm storage of well-described linguistic records is crucial to language documentation.
- Archives do not have the resources to prepare data.
- Training is essential for the integration of an archival sensibility into a linguist's fieldwork methods.
- It is up to the fieldworker to produce archival material from their fieldwork.