Reusing manuscript vocabularies

Nick Thieberger
University of Melbourne
ICLDC3, March 2\textsuperscript{nd}, 2013
The problem

- How to make thousands of pages of manuscript vocabularies accessible for ongoing reuse?
- How to annotate them for discovery?
  - Metadata describing their contents
- How to transcribe them?
Example of the Capell papers

- 14,000 pages of notes
- Sufficient metadata to locate the pages
- No resources to transcribe them
- Far more accessible than paper
Example of the Capell papers

- 14,000 pages of notes
- Sufficient metadata to locate the pages
- No resources to transcribe them
- Far more accessible than paper


Arthur Capell (1902-1986), Guide to Records

Listed by Peter Newton, Austhec and PARADISEC

Arthur Capell was an Australian linguist and ethnographer who spent much time recording and documenting both Australian Aboriginal languages and endangered languages in the Asia - Pacific region. The Arthur Capell textual collection consists of some thirty boxes of notes, transcripts, and other supporting materials. What is documented in this guide are some of the records relating to his non-Australian linguistic work that have been digitally imaged.

Please do not copy material from this site for further distribution but rather link to this site. PARADISEC has raised funds to digitise this collection and would like to be recognised for the work that we have put into developing the online presentation of fieldnotes. If you copy and distribute this data and do not acknowledge PARADISEC’s work then we will have to put password protection on the data.

Copyright: Paradisec believes that many of the items provided through this guide are no longer the subject of copyright restrictions, or have been cleared for display in this service by the Copyright owners. However, Paradisec invites any individuals who believe they hold current rights over items provided through this service to make contact.

[Details]

- About the records
  - Scope and content
  - How to use this finding aid
  - Archival terms

- Provenance - creators and former custodians

- Series list and summary descriptions

- Inventory listing by series
  - Series 01 - Personal and Biographical Material
  - Series 02 - General Linguistic and Ethnological Materials
  - Series 03 - Indonesia - Regions and Languages
  - Series 04 - Bougainville, Buka and Offshore Islands - Regions and Languages
  - Series 05 - Melanesia and Oceania - Regions and Languages
  - Series 06 - Solomon Islands - Regions and Languages
  - Series 07 - Timor - Regions and Languages
Example of the Capell papers

14,000 pages of notes
Sufficient metadata to locate the pages
No resources to transcribe them
Far more accessible than paper

Example of the Capell papers

Arthur Capell (1902-1986), Guide to Records

Arthur Capell was an anthropologist documenting both the Australian and Melanesian regions. The Arthur Capell papers contain notes and supporting materials for his Australian linguistic work.

Please do not copy.

Related Series

Part of

06 Solomon Islands

Inventory Listing

AC2-SIGELA101 A grammar of Gela (Ngelina) by C.H. Brooke and Melanesian Language. Ivens did not conduct field work in this language.

Creator Capell, Arthur

Date range 1937 Quantity 0.19 cm, 1 folio

Inventory Identifier SIGELA101 Series SIGELA

Series 06 - Solomon Islands - Regions and Languages

Series 07 - Timor - Regions and Languages

Example of the Capell papers

14,000 pages of notes
Sufficient metadata to locate the pages
No resources to transcribe them
Far more accessible than paper

Daisy Bates vocabularies

• 24,700 page images
• Mixed types of material
  – Handwritten manuscripts
    • Questionnaires (based on a template) (13,487 pp)
    • Freeform lists (1,251 manuscript pages)
  – Typescripts
    • Of questionnaires (4,933 pp)
    • Of freeform lists (some related to handwritten forms, but some not)
### Bates collection

**Summary of Sections and Folios**  
(91 boxes)

<table>
<thead>
<tr>
<th>Section</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>I Origins</td>
<td>1 - 2</td>
</tr>
<tr>
<td>II Geographical distribution</td>
<td>3 - 4</td>
</tr>
<tr>
<td>III Social organisation</td>
<td>5 - 14</td>
</tr>
<tr>
<td>IV Initiation ceremonies</td>
<td>15 - 16</td>
</tr>
<tr>
<td>V Totems</td>
<td>17 - 20</td>
</tr>
<tr>
<td>VI Religious beliefs, superstitions</td>
<td>21 - 23</td>
</tr>
<tr>
<td>VII Myths and legends</td>
<td>24 - 26</td>
</tr>
<tr>
<td>VIII Native food</td>
<td>27 - 28</td>
</tr>
<tr>
<td>IX Weapons, arts, crafts</td>
<td>29 - 31</td>
</tr>
<tr>
<td>X Diseases, remedies, death, burial</td>
<td>32 - 33</td>
</tr>
<tr>
<td>XI Dances, songs</td>
<td>34 - 36</td>
</tr>
<tr>
<td>XII Language: grammar and vocabularies</td>
<td>37 - 61</td>
</tr>
<tr>
<td>XIII General notes on Aborigines</td>
<td>62 - 66</td>
</tr>
</tbody>
</table>

“The book I edited for publication, *The Native Tribes of Western Australia*, contains only a fraction from the rich mine of information Daisy Bates discovered in her long years living with the Aborigines in Western Australia and South Australia. There is plenty of research still to do on the Bates Collection.”

(White 1993: 57)

500 copies of a blank wordlist questionnaire produced by the Government Printer in Perth in 1904

Authorised by the WA Protector, Malcolm A.C. Fraser

Around 220 filled out
Instructions for filling in particulars concerning the Language, Customs, and Habits of the Aboriginal Natives of Western Australia.

It is particularly requested that each person who courteously undertakes to render assistance in this work by supplying the required information will, on pages 3 to 75, insert, opposite the English words appearing there, the equivalent word or words in the dialect spoken by the Aboriginal Natives of the immediate neighbourhood. On pages 77 to 85 should be supplied native expressions having the same meaning as the short sentences there given, and on pages 87 to 97 replies, as complete as possible, should be furnished to the questions therein contained.

The heads into which the subjects of inquiry have been divided are as follows:

<table>
<thead>
<tr>
<th>Vocabulary</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Man, his Relationships, etc.</td>
<td>3 - 75</td>
</tr>
<tr>
<td>Parts and functions of the body</td>
<td>5 - 9</td>
</tr>
<tr>
<td>Animals</td>
<td>11 - 13</td>
</tr>
<tr>
<td>Birds</td>
<td>15 - 22</td>
</tr>
<tr>
<td>Fishes</td>
<td>23 - 24</td>
</tr>
<tr>
<td>Reptiles</td>
<td>25</td>
</tr>
<tr>
<td>Insects</td>
<td>27 - 28</td>
</tr>
<tr>
<td>The elements, etc.</td>
<td>29 - 36</td>
</tr>
<tr>
<td>General vocabulary</td>
<td>87 - 75</td>
</tr>
<tr>
<td>Short Sentences...</td>
<td>77 - 85</td>
</tr>
<tr>
<td>Questions</td>
<td>87 - 97</td>
</tr>
</tbody>
</table>

In obtaining the native equivalent of the English word, very great care should be taken to ascertain exactly how it is pronounced by the native, and then, in writing it down, it should be spelt phonetically, that is, to represent as nearly as practicable the exact sound made use of. Each word should be divided into syllables as it is naturally divided in the pronunciation, and an accent should be placed over the syllable or syllables on which emphasis or stress is laid, as, for instance, in the Murchison word for “Plain”—“Yal’-ga.”

Natives, as a rule, have distinct names for every minute portion of the human frame and other natural objects, and consequently very great care should be exercised to obtain the exact native equivalent of each word appearing in the list.

Thus, in endeavouring to obtain the native equivalent, say, of the right arm, it is possible, unless great care is taken, that the native word for hand, wrist, fore-arm, elbow, or upper arm of
In obtaining the native equivalent of the English word, very great care should be taken to ascertain exactly how it is pronounced by the native, and then, in writing it down, it should be spelt phonetically, that is, to represent as nearly as practicable the exact sound made use of. Each word should be divided into syllables as it is naturally divided in the pronunciation, and an accent should be placed over the syllable or syllables on which emphasis or stress is laid, as, for instance, in the Murchison word for “Plain”—“Yal’-ga.”
The following System of Orthography for Native Names of Places adopted by the Council of the Royal Geographical Society, the Foreign and Colonial Offices, Admiralty, and War Office, is to be adhered to in all Intelligence Division Publications.

1. No change is made in the orthography of foreign names in countries which use Roman letters; thus Spanish, Portuguese, Dutch, etc., names will be spelt as by the respective nations.

2. Neither is change made in the spelling of such names in languages which are not written in Roman character as have become by long usage familiar to English readers: thus Calcutta, Cairo, Calais, Moscow, etc., will be retained in their present form.

3. The true sound of the word as locally pronounced will be taken as the basis of the spelling.

4. An approximation, however, to the sound is alone aimed at. A system which would attempt to represent the more delicate inflections of sound and accent would be so complicated as only to defeat itself. Those who desire a more accurate pronunciation of the written name must learn it on the spot by a study of local accent and pronunciation.

5. The broad features of the system are:
   (a) That vowels are pronounced as in Italian and consonants as in English.
   (b) Every letter is pronounced, and no redundant letters are introduced. When two vowels come together, each one is sounded, though the result, when spoken quickly, is sometimes smeared to be distinguished from a single sound, as in ai, ou, ei.
   (c) One accent only is used, the acute, to denote the syllable on which stress is laid.

6. Indian names are accepted as spelt in Hunter’s Gazetteer of India, 1881.

The following amplification of these rules explains their application:

<table>
<thead>
<tr>
<th>Letters</th>
<th>Pronunciation and Remarks</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>is always soft, but it is so nearly the sound of e that it should be seldom used.</td>
<td>Calcutta.</td>
</tr>
<tr>
<td>b</td>
<td>is always soft as in church</td>
<td>China.</td>
</tr>
<tr>
<td>c</td>
<td>English d.</td>
<td>Ching-chin.</td>
</tr>
<tr>
<td>d</td>
<td>English j. sh should not be used for the sound of j. Thus, not Kajikumby, but is always hard. (Soft j is given by y)</td>
<td>Hailiang, Naf. Shangha.</td>
</tr>
<tr>
<td>e</td>
<td>is always pronounced when inserted.</td>
<td>Galápagos.</td>
</tr>
<tr>
<td>f</td>
<td>English j. Dy should never be put for this sound.</td>
<td>Hwang ho, Ngan hwi.</td>
</tr>
<tr>
<td>g</td>
<td>English k. It should always be put for the hard g.</td>
<td>Japan, Jinchuen.</td>
</tr>
<tr>
<td>h</td>
<td>The Oriental guttural...</td>
<td>Korea.</td>
</tr>
<tr>
<td>i</td>
<td>is another guttural, as in the Turkish...</td>
<td>Khmen.</td>
</tr>
<tr>
<td>j</td>
<td>As in English.</td>
<td>Kinh.</td>
</tr>
<tr>
<td>k</td>
<td>As in English.</td>
<td>K compliant, Mokpo.</td>
</tr>
<tr>
<td>l</td>
<td>As in English.</td>
<td>Bethlehlem.</td>
</tr>
<tr>
<td>m</td>
<td>As in English.</td>
<td>Kwangtung.</td>
</tr>
<tr>
<td>n</td>
<td>As in English.</td>
<td>As in English.</td>
</tr>
<tr>
<td>o</td>
<td>always a consonant, as in good, and therefore should never be used as a terminal, t or c being substituted as the sound may require.</td>
<td>Kikyo.</td>
</tr>
<tr>
<td>p</td>
<td>As in English.</td>
<td>Mikidomi.</td>
</tr>
<tr>
<td>q</td>
<td>As in English.</td>
<td>Kwali.</td>
</tr>
<tr>
<td>r</td>
<td>As in English.</td>
<td>Zola.</td>
</tr>
<tr>
<td>s</td>
<td>As in English.</td>
<td>Mombasa.</td>
</tr>
<tr>
<td>t</td>
<td>The French j, or s in treasure...</td>
<td>Tangpaia, Galápagos, Palawan, Sarawak.</td>
</tr>
<tr>
<td>u</td>
<td>Assents should not generally be used, but where there is a very decided emphatic syllable or stress, which affects the sound of the word, it should be marked by an acute accent.</td>
<td></td>
</tr>
</tbody>
</table>

N.B.—These rules were revised and approved by the Council, Royal Geographical Society, 11th December, 1901.
Manuscript vocabularies

- Filled out by a number of respondents
- Varying approaches to orthographies
- Subsequently typed under Bates’s supervision
General Vocabulary (continued)

Kine (my own) Janna jerkæ, ngai janna, that is mine.

Koeck, to Yinn wadia war

Monster (fabulous) Jeeæ, tochorra, ngaongororo

Music Jirim ñfrin, lererr
Beombul-doumbul = marking time by striking ground with the palms of the hands.)

My Ngai Janna

Baked Køya

Bane Eyjelwel

Barrow Jingin

Native well Bélarra

Bear Jammara

Beaver Arressanga, maaleo

Beech Marsoo, laip

Be No maaleo, arressanga, arre

Boby Arressanga

Boiss Inganga

Bone Maaleo, arressanga

Booh: Wardeo, waamoe

Bone-piercing Warraajin nimam

Bone-stick Karimba, lerria

Bot Maaleo

Bow (at once) Mamungara

Offend Hoberduk, joojerdouk

Oh! Warrai

Old Jabil

One Maajeroe

Ours Jarrarda, jarrods
Man, his relationships, etc. (cont.)

Sister
Waldulu, d'Angulu, kabal

Sister-in-law
D'amirba, wald'urunia, d'ugabu, yoga

Son
Wandi

Son-in-law
Wandi

Cormorant
Yibirdijalu

Diver (sea)
Dhulia (also penguin)

Duck (generic)
Bundu kadea

Ngurarba

Murdi-ning or Murding. m. 'murderer' (bird man)

Jirêlo, marain maraín, nyijí waran gülbarding

According to Henry (Karnduing), the yimma is = kundain walja
kurdudu and mugula
heart and belly

Walja miru = pointers (star)
Eaglehawk's miro

Yinma wálga, yimma markings

Bulburda, m., a Malgoram at Eucla had karongu and other things.
Section XII - Vocabularies

Publicly available since the 1950s (acquired by the National Library of Australia in 1941)

Used in Native Title cases, teaching resources, comparative work
Members of Aboriginal communities are advised that this listing contains names of deceased persons. Some titles have been transcribed from original items and reflect attitudes of the time. Daisy Bates (1859 – 1951) was a pioneer in the observation, over a period of 35 years, of the Aboriginal people living in the desert around the Great Australian Bight. Born in County Tipperary, Ireland as Daisy May O’Dwyer, she arrived in Australia in 1883 and worked as a governess to the Bates family near Nowra, NSW. In 1885 she married the eldest son, drover John Bates. It has since been proven that she had previously married the man we know as Harry ‘Breaker’ Morant and possibly another man as well, without ever having had these marriages dissolved. After bearing a son to John Bates, she left alone for London to work as a journalist for the next 5 years. Daisy Bates returned to Australia in 1899 and consulted with Aboriginal people on language, religion, myths, customs and kinship. In about 1916 she settled in Ooldea SA to live with her people for more than 16 years, seeing her role as ‘smoothing the pillow of a dying race’. She wrote My Natives and I and The Passing of the Aborigines in 1935 in order to publicise her cause and attract funds for her work. Some of her notes were compiled with secretarial help. 1936 – 40 and later deposited in the National Library Sources ADB, Obituary

Sub-communities within this community

- Series 12: Manuscript of ‘Native tribes of Western Australia’ [13]

- Series 7: Myths and Legends [1]

- Series 9: Photographs [9]
NLA microfilms

• All manuscripts were microfilmed some time ago (55 reels, of which Section XII represents 15 reels)

• Arts Faculty at the University of Melbourne paid for these to be digitised
Naming

The naming convention builds on the identifiers used by the NLA:

- folio – page number (Typescripts – T) Part number
  - 51-007T-XII2D15.tif

- folio – booklet & page number for questionnaires (Q)
  - 39-241Q99-XII2A5.tif

- folio – page number for other manuscripts (M)
  - 39-308M-XII2A.tif
List of (122) individual speakers whose names are included in the catalog:

| Baaburgurt | Boonyooy | Jooloor | Mundhongo | Waienburt |
| Baandee    | Boordenam | Joowel | Ngadain | Wajida    |
| Bailberin  | Booreejee | Jubyah | Ngalbaitch | Wallbarring |
| Baiungan   | Bulgara | Jukum Nanga | Ngalloodha | Wanbojee |
| Balbuk     | Burangani | Kaiar | Ngalyart | Wardabirt |
| Balgalea   | Burdali | Kajaman | Ngilgee | Wardunda |
| Balgomain  | Deebungool | Kajilgarra | Ngindilha | Wardunya |
| Bandila    | Dharroon | Kalagurda | Ngoonbib | Warndan |
| Banyaitch  | Dhoonguljarra | Karnduing | Ngoonemurr | Werdbart |
| Bardeet    | Dhoornée | Karree | Ngoonguea | Wering |
| Bardil     | Gauera | Katamina | Ngoonjicja | Wildeejinna |
| Beedagooroo | Goolgool | Katurnumia | Ngunaipon | Willambung |
| Beedee     | Goolaara | Kokobina | Ngwoonbib | Wingadee |
| Beejee     | Gunmaguru | Kokobung | Nimbool | Wirijan |
| Beeleenan  | Gweea | Kondeebomur | Nindilya | Woollberr |
| Bijarda    | Ibari | Kuliji | Notum | Woogomirt |
| Bilbil     | Ijala | Kumandura | Nyarda nyardareej | Yabbaroo |
| Biljarda   | Ilaurrainya | Majjere | Nyaul | Yagganguri |
| Billingeec | Indar | Manya | Nyeerbeeji | Yalguru |
| Bindaweereec | Injiri | Marratharra | Nyimbaija | Yoloyn |
| Binilya    | Jaal | Minbunga | Roorrje | Yoolbaree |
| Binyunjoo  | Jakbun | Minjia | Turada | Yoolbarga |
| Boolyahung | Jimmer | Monnop | Wabbingan | Yowinbarga |
| Boongong   | Jinguru | Muggabeen | Wabbinyet |  |
| Boonjerin  | Joobaitch | Mulgarongu | Wadunda |  |
SECTION XII

LANGUAGE : GRAMMAR AND VOCABULARIES

Names given are of natives from which vocabularies were obtained.

Part 2. E. 1.(c)

Northwestern District

Jukam Nanga of Broome.

Folio 52/105-131
The nature of the lists

• Typically a word and equivalent
• Occasionally
  – Additional words
  – Example sentences
  – Extra paragraphs or pages of information
Problems

Words with >1 equivalent

Dead  jingah = dead, mnedu = about to die
      mango = dead person, mimebo gadding
      yowie = just now dead
      coombah = by and by dead
      yennal = dead and leave him behind
      and blackfellow run away
Problems

Words with >1 equivalent

Light
chanera, ngorla

‘ngorla’ – light weight
‘chanera’ – light (not dark)
Problems

• Variation in:
  – Orthographies
  – Legibility of handwriting
  – Coverage per list
OCR or typing?

- Image to text (microfilm to TIF)
  - OCR – low success rate

- keyboarding text with some TEI codes
Bates and publicly funded research

- Large scale, publicly funded comparative project on Australian languages begun in the 1970s
  - Analysis is published
  - Primary material, some of which analyses Daisy Bates’s work, is not published
  - No tradition at the time of preparing primary data for re-use
Bates and publicly funded research

• Native Title research
  – No central repository
  – No standard approach to citation or presentation of primary data
Presentation of primary data

e.g., Laves project

As reported in:
Text Encoding Initiative (TEI)

- TEI established in 1980s
- Revised versions (now version 5)
- Explicit encoding of elements of the text

http://www.tei-c.org/index.xml
Why use the TEI?

• Create once, use many times
  – Various outputs
    • Comparative database
    • Epub, HTML, Word processor documents

• Community of TEI users
  – Tools
  – Methods
Neatline

Neatline is a mix-and-match toolset for the creation of interlinked timelines and maps as interpretive expressions of the literary or historical content of archival collections. It allows scholars and archivists to build on standard descriptive metadata and georectified maps to produce rich, evocative – even theoretical – spatial and temporal visualizations of the textual content of catalogued letters, manuscripts, and artifacts.

Neatline is therefore a geo-temporal framework for fruitful interchange among scholars and the stewards of primary resources. It builds on Omeka, OpenLayers, GeoServer, and SIMILE Timeline. Adam Soroka and
First step

- Keyboarding in structured format
  - XML template
- “Double key, double compare”
- Encoding for simple content
  - <add> handwritten or extra information
Fish (generic) n-ar-da
Oyster (large Pearl) <add>Oyster
Frog (generic) ma-lee
Iguana, long-tailed wongalee
Lizard (generic) boon-jar-da
Snake (generic) warl-la
Ant (common small) bing-a
Ant (generic) biyu
Ant (white) muyo
Afternoon beer-inga
Cold mool-tha
Country (open) bulgarra
Country (story) munda-manda
Creek moorie
Dark <add>Dark
Floods cur-ry-aca
Grass min-na
Hole mar-do
Moon weel-arr
Morning wunnunda
Pool in River babba wyu
Rain yoong-go
River wandi
<TEI xmlns:tei="http://www.tei-c.org/ns/1.0" xmlns="http://www.tei-c.org/ns/1.0">
  <!-- add page numbers
  Include TS page numbers as <pb facs=""/> manuscript pages given as pb n=""
  Include additions that are not in the current template as <add> see entry 0003a below-->
  <teiHeader>
    <fileDesc>
      <titleStmt>
        <title>Daisy Bates Vocabulary</title>
      </titleStmt>
      <publicationStmt>
        <availability>
          <p>NLA MS 365</p>
        </availability>
      </publicationStmt>
      <sourceDesc>
        <listPerson>
          <person>
            <persName>[This is where the named authors of the document should be listed]</persName>
          </person>
        </listPerson>
        <listPlace>
          <place>
            <region>[This is where the place named in the source should be provided]</region>
          </place>
        </listPlace>
      </sourceDesc>
    </fileDesc>
    <text>
      <body>
        <div>
          <pb n="3" type="page"/>
          <head>Man, his relationships, etc.</head>
          <p><gloss n="0001" xml:lang="en">Aunt</gloss><term xml:lang="XXX">XXXX</term></p>
          <p><gloss n="0002" xml:lang="en">Baby</gloss><term xml:lang="XXX">XXX</term></p>
          <p><gloss n="0003" xml:lang="en">Blackfellow</gloss><term xml:lang="XXX">XXX</term></p>
        </div>
      </body>
    </text>
  </teiHeader>
</TEI>
Current state of the pilot project

- All images digitised and renamed
- Metadata system in place
- ~1431 pages keyboarded and encoded in a TEI-conformant form
- Wordlists geocoded and mapped
<table>
<thead>
<tr>
<th>Title</th>
<th>Location</th>
<th>Lat</th>
<th>Long</th>
<th>Language</th>
<th>Notes</th>
<th>Total keyboard pages</th>
<th>Q</th>
<th>M</th>
<th>No Q</th>
<th>No M</th>
<th>Nocmage numbers</th>
<th>Microfilm number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vocabulary from Albany &amp; Denmark, folio</td>
<td>Albany</td>
<td>34.07630</td>
<td>117.91482</td>
<td>Nyungar</td>
<td>yes</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mfm 026 337</td>
</tr>
<tr>
<td>Boerendam, Be Quarterly Wardat and Boonging of Albany, folio</td>
<td>Albany</td>
<td>35.02382</td>
<td>117.88472</td>
<td>Nyungar</td>
<td>yes</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mfm 026 337</td>
</tr>
<tr>
<td>Woongorri of Albany, folio</td>
<td>Albany</td>
<td>35.02062</td>
<td>117.88472</td>
<td>Nyungar</td>
<td>yes</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mfm 026 337</td>
</tr>
<tr>
<td>Bartet of Bridgetown, folio</td>
<td>Bridgetown</td>
<td>33.95704</td>
<td>116.13723</td>
<td>Nyungar</td>
<td>yes</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mfm 026 337</td>
</tr>
<tr>
<td>Ngapilak of Karruk, folio</td>
<td>Bridgetown</td>
<td>33.95704</td>
<td>116.13723</td>
<td>Nyungar</td>
<td>yes</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mfm 026 337</td>
</tr>
<tr>
<td>Ngapilak of Jerramungia, folio</td>
<td>Jerramungia</td>
<td>33.94238</td>
<td>118.92151</td>
<td>Nyungar</td>
<td>yes</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mfm 026 337</td>
</tr>
<tr>
<td>Ngapilak of Esperance, MSS, folio</td>
<td>Esperance</td>
<td>33.85941</td>
<td>121.39324</td>
<td>Nyungar</td>
<td>yes</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mfm 026 337</td>
</tr>
<tr>
<td>Ngapilak of Esperance, MSS, folio</td>
<td>Esperance</td>
<td>33.85941</td>
<td>121.39324</td>
<td>Nyungar</td>
<td>yes</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mfm 026 337</td>
</tr>
<tr>
<td>[2] Shympy of Ravenshorpe, MSS, folio</td>
<td>Ravenshorpe</td>
<td>33.58161</td>
<td>120.04160</td>
<td>Nyungar</td>
<td>yes</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mfm 026 337</td>
</tr>
<tr>
<td>[4] Boerendam, Be Quarterly Wardat and Boonging of Albany, MSS, folio</td>
<td>Albany</td>
<td>34.07680</td>
<td>117.91482</td>
<td>Nyungar</td>
<td>yes</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mfm 026 337</td>
</tr>
<tr>
<td>Miscellaneous</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mfm 026 337</td>
</tr>
</tbody>
</table>
Current state of the pilot project

• Design of database and presentation platform under way

• Protocols for dissemination?
  – already published for >50 years
  – NLA and Barr Smith Library have already worked out dissemination
  – Is web-based dissemination any different?
Outcomes

- First application of TEI to an Australian language corpus
  - creation of an archival textual version
- Linking textual version of manuscript to image of manuscript
- Online annotation of text based on images of manuscripts (possibility of crowdsourcing = having volunteers do the work)
A note from the past

This is a piece of text discovered in Egypt, written over 1000 years ago.

We'd like you to help us to read it.

Images may not be copied or offloaded, and the images and their texts may not be published. All digital images of the Oxyrhynchus Papyri are © Imaging Papyri Project, University of Oxford. The papyri themselves are owned by the Egypt Exploration Society, London. All rights reserved.

http://ancientlives.org/tutorial/transcribe
Outcomes

• Making textual versions of the vocabularies accessible
• Increasing visibility of (tagged) elements of the files (personal names, placenames) and potential to link to other datasets via an API
• Comparative (geocoded) wordlists
• Collaboration with National Library of Australia and Barr Smith Library
• Social outcomes – providing access to archival material for descendants of speakers

Thanks to research assistants Lauren Gawne and Jill Vaughan

Funded by the Faculty of Arts at the University of Melbourne and ARC grant DP0984419 – Doing great things with small languages: Safeguarding Indigenous language material of Australia's region by clever use of new technology. CIs Nick Thieberger and Rachel Nordlinger