Micro Data Repositories: Increasing the Value of Research on the Web

Adam Field, Patrick McSweeney, iSolutions, University of Southampton
"I suppose it is tempting, if the only tool you have is a hammer, to treat everything as if it were a nail.”

-- Abraham Maslow
“Does it Fit?  Will it Last?  How Much?  We Need a Project!”  -- Computing Services
“Look What I Made!”

-- Hetty Field
“Can we put it in a Standard System?”

-- Computing Services

• Institutional Repository
  – Treat data as a publication

• Online Repository
  – youtube / figshare / dropbox / github

• Bespoke Solution
  – One-off BD-driven WebApp
  – Static HTML Site

• Dataset Repository
  – e.g. ecrystals, Linnean Collection
Micro Data Repositories

“A default repository that has undergone a minimal set of changes to turn it into a showcase for a dataset.” -- Adam Field
Case Study - MedMus

• Webification of reference book data on precise spelling of medieval French Refrains
• Complex data model (refrains appear in works which may appear in other works which are written in manuscripts)
• 18 months and not finished yet (see next slide)
“What’s a Data Model?” -- Music Department Professor

“All the Above” -- Adam Field
Case Study – LangSnap

• Iterative assessment of non-native language learners before, during and after a year-long residence in the country of the language.
• Transcripts and recordings (MP3 and WAV)
• Previous projects on hand-build websites
• Good, clear understanding of the data model
Emergency Screen Shots

French Learner Language Oral Corpora

Home  About us  Team members  Ground rules

The FLLOC project

Project description
Rationale
CHILDES
Additional conventions
Flloc XML
Description of corpora
Download Data
Search the corpora

Year 12 Modern Times Task

View all datasets for the Newcastle Corpus

<table>
<thead>
<tr>
<th>Wave File</th>
<th>MP3 File</th>
<th>Transcription File</th>
<th>Tagged File</th>
<th>XML File</th>
</tr>
</thead>
<tbody>
<tr>
<td>T01JUL12</td>
<td>T01JUL12</td>
<td>T01JUL12</td>
<td>T01JUL12</td>
<td>T01JUL12</td>
</tr>
<tr>
<td>T02OMA12</td>
<td>T02OMA12</td>
<td>T02OMA12</td>
<td>T02OMA12</td>
<td>T02OMA12</td>
</tr>
<tr>
<td>T04VIV12</td>
<td>T04VIV12</td>
<td>T04VIV12</td>
<td>T04VIV12</td>
<td>T04VIV12</td>
</tr>
<tr>
<td>T05ANA12</td>
<td>T05ANA12</td>
<td>T05ANA12</td>
<td>T05ANA12</td>
<td>T05ANA12</td>
</tr>
<tr>
<td>T08ERI12</td>
<td>T08ERI12</td>
<td>T08ERI12</td>
<td>T08ERI12</td>
<td>T08ERI12</td>
</tr>
<tr>
<td>T10VIV12</td>
<td>T10VIV12</td>
<td>T10VIV12</td>
<td>T10VIV12</td>
<td>T10VIV12</td>
</tr>
<tr>
<td>T11VIV12</td>
<td>T11VIV12</td>
<td>T11VIV12</td>
<td>T11VIV12</td>
<td>T11VIV12</td>
</tr>
<tr>
<td>T12VIV12</td>
<td>T12VIV12</td>
<td>T12VIV12</td>
<td>T12VIV12</td>
<td>T12VIV12</td>
</tr>
<tr>
<td>T13ERI12</td>
<td>T13ERI12</td>
<td>T13ERI12</td>
<td>T13ERI12</td>
<td>T13ERI12</td>
</tr>
<tr>
<td>T14ANA12</td>
<td>T14ANA12</td>
<td>T14ANA12</td>
<td>T14ANA12</td>
<td>T14ANA12</td>
</tr>
<tr>
<td>T15ANA12</td>
<td>T15ANA12</td>
<td>T15ANA12</td>
<td>T15ANA12</td>
<td>T15ANA12</td>
</tr>
<tr>
<td>T16JUL12</td>
<td>T16JUL12</td>
<td>T16JUL12</td>
<td>T16JUL12</td>
<td>T16JUL12</td>
</tr>
<tr>
<td>T17VIV12</td>
<td>T17VIV12</td>
<td>T17VIV12</td>
<td>T17VIV12</td>
<td>T17VIV12</td>
</tr>
<tr>
<td>T18OMA12</td>
<td>T18OMA12</td>
<td>T18OMA12</td>
<td>T18OMA12</td>
<td>T18OMA12</td>
</tr>
<tr>
<td>T20ERI12</td>
<td>T20ERI12</td>
<td>T20ERI12</td>
<td>T20ERI12</td>
<td>T20ERI12</td>
</tr>
</tbody>
</table>
### Browse Items

Items may be browsed by the following:

- Language
- Participant
- Activity
- Collection Round

### Browse by Participant

<table>
<thead>
<tr>
<th>Language</th>
<th>Participant</th>
<th>Round</th>
<th>Activity</th>
<th>Files</th>
<th>Files Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spanish</td>
<td>169</td>
<td>Pretest</td>
<td>Interview</td>
<td>D169aNTV.mp3 D169aNTV.wav D169aNTV.cha</td>
<td>Full Record...</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>D169aNTV.mor.pst.cex</td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>169</td>
<td>Pretest</td>
<td>Narrative (Cat Story)</td>
<td>C169aNTV.mp3 C169aNTV.WAV C169aNTV.cha</td>
<td>Full Record...</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>C169aNTV.mor.pst.cex</td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>169</td>
<td>Pretest</td>
<td>Writing (Gay Adoption)</td>
<td>G169a.cha G169a.mor.pst.cex</td>
<td>Full Record...</td>
</tr>
<tr>
<td>Spanish</td>
<td>169</td>
<td>Visit Abroad 1</td>
<td>Interview</td>
<td>D169bNTV.mp3 D169bNTV.wav D169bNTV.cha</td>
<td>Full Record...</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>D169bNTV.mor.pst.cex</td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>169</td>
<td>Visit Abroad 1</td>
<td>Narrative (Sisters Story)</td>
<td>S169bNTV.mp3 S169bNTV.WAV S169bNTV.cha</td>
<td>Full Record...</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>S169bNTV.mor.pst.cex</td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>169</td>
<td>Visit Abroad 1</td>
<td>Writing (Drugs)</td>
<td>D169b.cha D169b.mor.pst.cex</td>
<td>Full Record...</td>
</tr>
</tbody>
</table>

**Up a level**

For more information, visit the [LangSnap Dev](http://www.langsnap.dev) website.
**Participant 169, Pretest, Interview**

<table>
<thead>
<tr>
<th>Information</th>
<th>Audio</th>
<th>Documents</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Language:</strong> Spanish</td>
<td>Audio (MP3)</td>
<td>Text (Transcript)</td>
</tr>
<tr>
<td><strong>Participant ID:</strong> 169</td>
<td>0169aNTV.mp3 - Audio Recording</td>
<td>0169aNTV.cha - Transcript</td>
</tr>
<tr>
<td><strong>Investigator:</strong> NTV</td>
<td>Download (18MB)</td>
<td>Download (16kB)</td>
</tr>
<tr>
<td><strong>Data Collection Round:</strong> Pretest</td>
<td>Audio (Uncompressed)</td>
<td>Text (Transcript with Grammar Analysis)</td>
</tr>
<tr>
<td><strong>Activity:</strong> Interview</td>
<td>0169aNTV.wav - Audio Recording</td>
<td>0169aNTV.mor.pst.cex - Transcript with Grammar Analysis</td>
</tr>
<tr>
<td><strong>URI:</strong> <a href="http://langsnap-dev.soton.ac.uk/id/eprint/1076">http://langsnap-dev.soton.ac.uk/id/eprint/1076</a></td>
<td>Download (223MB)</td>
<td>Download (40kB)</td>
</tr>
</tbody>
</table>
Browse Refrains and Works

Items may be browsed by the following:

Refrains and Works

- Refrains and Works

Refrains

- by Singer
- by Manuscript
- by Words in Manuscript Reading
- by Circumstance
- by Location

Works

- by Author
- by Manuscript
- by Text
- by Location

Browse by Refrains and Works

Please select a value to browse from the list below.

- Refrain by vdB index
- Indexed Song
- Unindexed Song
- Motet by title
- Motet Part by index
- Narrative by title

vdB 10: Adés sont ces sades brunetes les plus jolietes.

1: In F-Pn fr. 1581, fol. 50 ra

Details

- Manuscript: F-Pn fr. 1581, fol. 50 ra
- Singer: Brune la femme de l'ours
- Audience: Bran l'ours
- Function: Discours amoureux
- Circumstance: Entrée à Passe-Orguell (fête finale)
- Preceding Lyric: Chanter, Chant
- Meter: ABA''A'

Reading Text(s)

- Adés sont ces sades brunetes les plus jolietes.
vdB 10: Adés sont ces sades brunetes les plus jolietes.

**Refrain**

**Refrain ID:** 10  
**Abstract Text:** Adés sont ces sades brunetes les plus jolietes.

<table>
<thead>
<tr>
<th>Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>in F-Pn fr. 1581</td>
</tr>
<tr>
<td>in F-Pn fr. 25566</td>
</tr>
<tr>
<td>in F-Pn fr. 1593</td>
</tr>
<tr>
<td>in F-Pn fr. 372</td>
</tr>
<tr>
<td>in F-NO H 196</td>
</tr>
</tbody>
</table>

**Instance**

**Manuscript** adés sont ces sades brunetes les plus jolietes  
**Readings:**

**Singer:** Brune la femme de l'ours  
**Audience:** Brun l'ours  
**Function:** Discours amoureux  
**Circumstance:** Entrée à Passe-Orgueil (fête finale)  
**Preceeding Lyric:** Chanter, Chant  
**Musical Structure:** AB  
**Meter:** A'8A'5  
**Manuscript Collocation:** F-Pn fr. 372, fol. 52 vb  
**Parent Work:** Renart le Nouvel [v. 6880 ]

**Music**

```
\[\text{music notation}\]
```

a - dés sont ces sa - des bru - ne - tes les plus io - li - e - tes
Conclusions

• Researchers like micro repositories
  – Data is usable in the browser
  – CMS features put management in their hands
  – Repository features make correcting data trivial

• Central Computing Services like don’t hate micro repositories

• We’re developing a standard procedure for rapidly creating these using EPrints.