Micro Data Repositories : Increasing the Value of Research on the Web
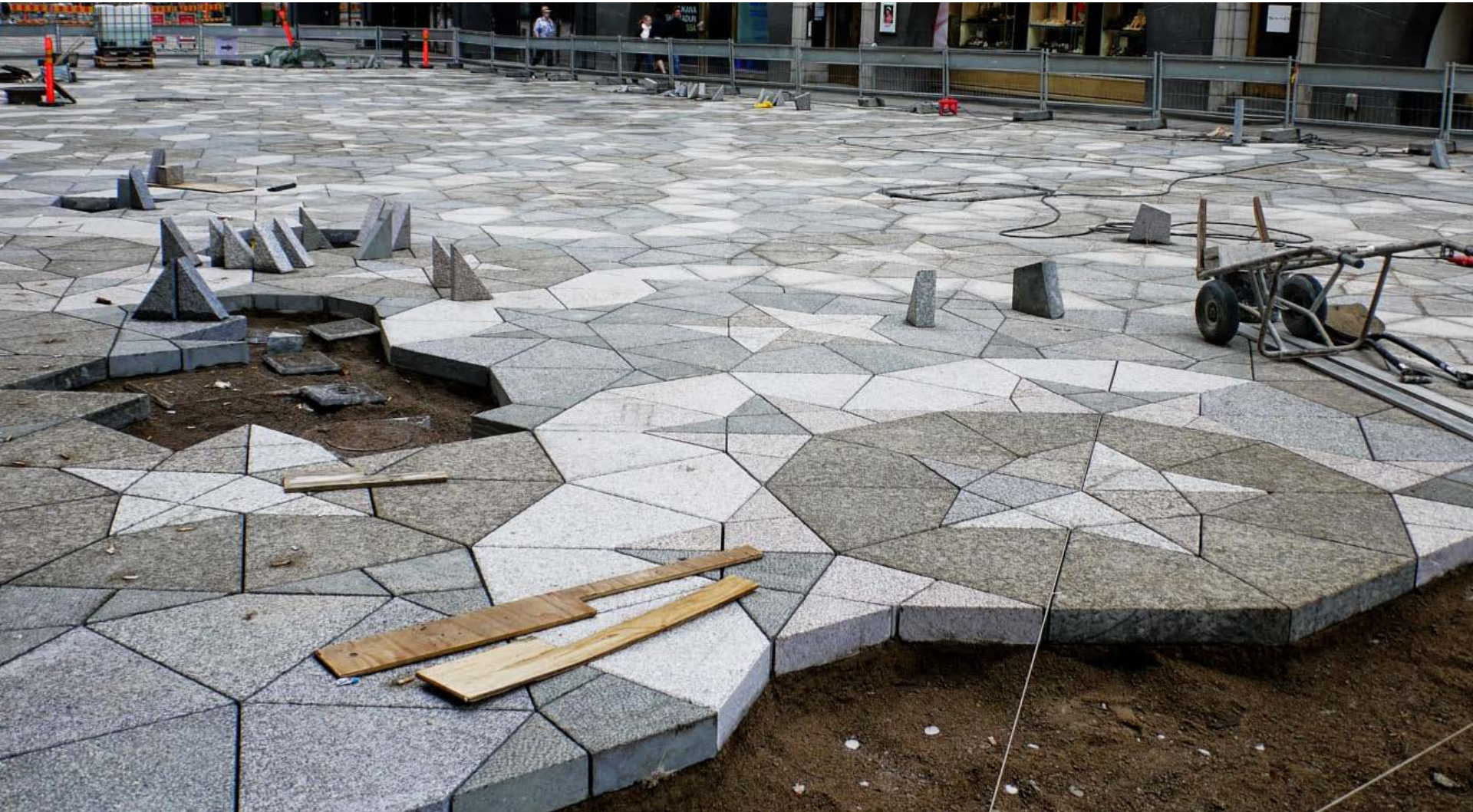
Adam Field, Patrick McSweeney,

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

CODE

"Does it Fit? Will is 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)
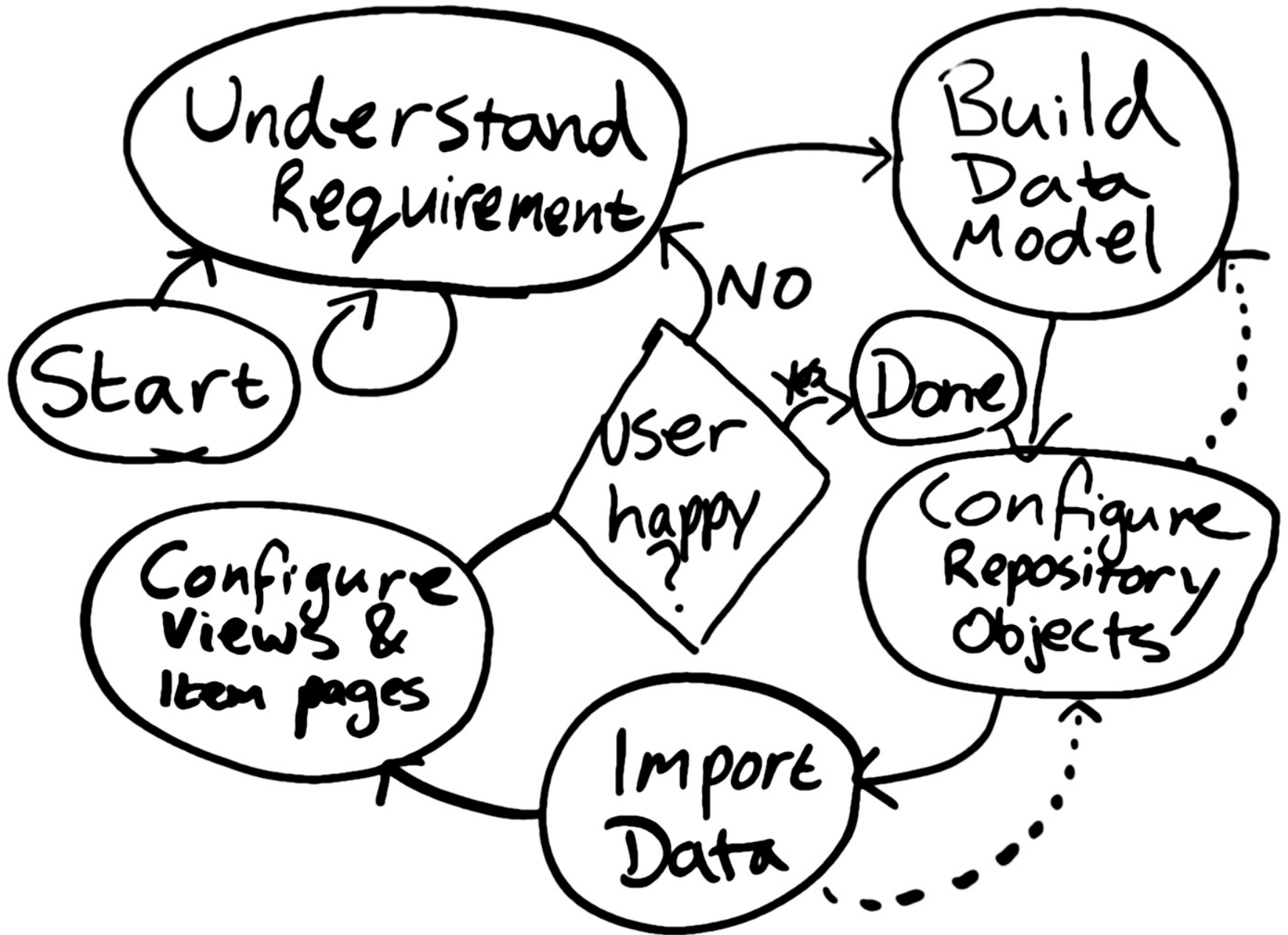
"What's a Data Model?"

-- Music Department Professor

"What's a Refrain?" -- 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

Start → Understand Requirement → Build Data Model → Configure Repository Objects → Import Data → Configure Views & Item pages → User happy? — NO → Understand Requirement / YES → Done

Live Demo

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

Corpora

LANGSNAP Corpus
Young Learners Corpus
Newcastle Corpus
Linguistic Development Corpus
Progression Corpus
Brussels Corpus
Reading Corpus
Salford Corpus
UEA Corpus

Year 12 Modern Times Task

View all datasets for the Newcastle Corpus

Wave File	MP3 File	Transcription File	Tagged File	XML File
T01JUL12	T01JUL12	T01JUL12	T01JUL12	T01JUL12
T02OMA12	T02OMA12	T02OMA12	T02OMA12	T02OMA12
T04VIV12	T04VIV12	T04VIV12	T04VIV12	T04VIV12
T05ANA12	T05ANA12	T05ANA12	T05ANA12	T05ANA12
T08ERI12	T08ERI12	T08ERI12	T08ERI12	T08ERI12
T10VIV12	T10VIV12	T10VIV12	T10VIV12	T10VIV12
T11VIV12	T11VIV12	T11VIV12	T11VIV12	T11VIV12
T12VIV12	T12VIV12	T12VIV12	T12VIV12	T12VIV12
T13ERI12	T13ERI12	T13ERI12	T13ERI12	T13ERI12
T14ANA12	T14ANA12	T14ANA12	T14ANA12	T14ANA12
T15ANA12	T15ANA12	T15ANA12	T15ANA12	T15ANA12
T16JUL12	T16JUL12	T16JUL12	T16JUL12	T16JUL12
T17VIV12	T17VIV12	T17VIV12	T17VIV12	T17VIV12
T18OMA12	T18OMA12	T18OMA12	T18OMA12	T18OMA12
T20ERI12	T20ERI12	T20ERI12	T20ERI12	T20ERI12

LangSnap Dev

University of **Southampton**

Login

Browse Items

Items may be browsed by the following:

- Language
- Participant
- Activity
- Collection Round

Browse by Participant

🔺 Up a level

Language	Participant	Round	Activity	Files	
Spanish	169	Pretest	Interview	O169aNTV.mp3 O169aNTV.wav O169aNTV.cha O169aNTV.mor.pst.cex	Full Record...
Spanish	169	Pretest	Narrative (Cat Story)	C169aNTV.mp3 C169aNTV.WAV C169aNTV.cha C169aNTV.mor.pst.cex	Full Record...
Spanish	169	Pretest	Writing (Gay Adoption)	G169a.cha G169a.mor.pst.cex	Full Record...
Spanish	169	Visit Abroad 1	Interview	O169bNTV.mp3 O169bNTV.wav O169bNTV.cha O169bNTV.mor.pst.cex	Full Record...
Spanish	169	Visit Abroad 1	Narrative (Sisters Story)	S169bNTV.mp3 S169bNTV.WAV S169bNTV.cha S169bNTV.mor.pst.cex	Full Record...
Spanish	169	Visit Abroad 1	Writing (Drugs)	D169b.cha D169b.mor.pst.cex	Full Record...

LangSnap Dev

Login

Participant 169, Pretest, Interview

Information

Language:	Spanish
Participant ID:	169
Investigator:	NTV
Data Collection Round:	Pretest
Activity:	Interview
URI:	http://langsnap-dev.soton.ac.uk/id/eprint/1076

Audio

Audio (MP3)
O169aNTV.mp3 - Audio Recording
Download (18MB)

Audio (Uncompressed)
O169aNTV.wav - Audio Recording
Download (223MB)

▶ ⬤──────────────── 0:00 🔇

Documents

Text (Transcript)
O169aNTV.cha - Transcript
Download (16kB)

Text (Transcript with Grammar Analysis)
O169aNTV.mor.pst.cex - Transcript with Grammar Analysis
Download (40kB)

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
- by

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 M index | Narrative by title |

- vdB 1: A bonne amour sui donnee mon vivant pour miex valoir. (1)
- vdB 2: A boine dame loiaus sui donnés. (2)
- vdB 3: A celi que je n'os nommer. (1)
- vdB 4: Acolez moi et baislez doucement, car li maus d'amer mi tient joliement. (6)
- vdB 5: A cuer me tient. (1)
- vdB 6: A cuer plain de joie, tout a son talent, servirai m'amie. (2)
- vdB 7: ...lement, me suix doneis sans retour, outreement. (1)
- ...ent com vivrai. (2)
- ...doloir. (3)
- ...s plus jolietes. (5)
- ...moi ne le sarés vous ja. (4)
- ...r je m'en vois souspirant en terre estrange. (3)
- ...qu'il les me gart. (16)
- ...nouvieles ai. (2)
- ...el lieu que vous soiés. (1)
- ...oi, fins cuers et plesant. (3)
- ...amors me tout solaz et ris. (7)
- ...s, pour estre miex amee d'amie fine. (1)
- ...onnees. (5)
- ...douné. (4)
- ...moi, qui me menez a mort si ne savez pour coi. (1)
- ...m'ochiés a tort? (2)

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: Brun l'ours
Function: Discours amoureux
Circumstance: Entrée à Passe-Orgueil (fête finale)
Preceeding Lyric: Chanter, Chant
Meter: A'8A'5

[Detailed View]

Reading Text(s)

- Adiés sont les sades brunetes les plus iolietes

Parent Work:

ID: N48/1
Authors: Jacquemart Gielée
Edition: Roussel (H.), Renart le nouvel par Jaquemart Gielee, Paris, A. et J. Picard et Cie, 1961
Generic Descriptor: Roman
Dating: Ca 1289
Primary Language: Langue d'oïl
Style of Discourse: Vers, Prose

[Detailed View]

2: In F-Pn fr. 25566, fol. 167 vb

Details

Manuscript: F-Pn fr. 25566, fol. 167 vb
Singer: Brune la femme de l'ours
Audience: Brun l'ours
Function: Discours amoureux
Circumstance: Entrée à Passe-Orgueil (fête finale)

Reading Text(s)

- Adés sont ces sades brunetes les plus iolietes.

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

Refrain

Refrain ID: 10

Abstract Text: Adés sont ces sades brunetes les plus jolietes.

Instances:
in F-Pn fr. 1581
in F-Pn fr. 25566
in F-Pn fr. 1593
in F-Pn fr. 372
in F-MO H 196

Instance

Manuscript Readings: adés sont ces sades brunetes les plus iolietes

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

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.