

# FCS at UPF

April 2013 Copenhagen

Marta Villegas

Requirement: find all Catalan books  
translated from Latin or Greek

“SRU + subcorcups indexing”

# MARC record from Catalan Library

..  
041 0 cat|alat → Source/target languages  
080 871-1"-00"  
100 1 Catul, Gai Valeri, |dca. 84-ca. 54 aC → Author  
240 10 Poesia. |lCatalà. |kSeleccions  
245 10 Poemes / |cCatul ; traducció i notes de Josep Ignasi → Title + more info  
Ciruelo & Jaume Juan  
260 Barcelona : |bEdhasa, |c1982  
300 XVIII, 254 p. ; |c16 cm  
490 1 La Garba ; |v2  
504 Índexs  
546 Text en català i llatí  
700 1 Ciruelo, José Ignacio, |etrad. → Translators  
700 1 Juan, Jaume, |etrad.  
710 2 Fons Néstor Luján (Biblioteca de Catalunya) |5ES-BaBC  
...

## DC record

- **Autor** [Catul, Gai Valeri, ca. 84-ca. 54 aC](#)
- **Títol uniforme** [Poesia. Català. Seleccions](#)
- **Títol**  
[Poemes / Catul ; traducció i notes de Josep Ignasi Ciruelo & Jaume Juan](#)
- **Publicació** Barcelona : Edhasa, 1982
- **Descripció** XVIII, 254 p. ; 16 cm
- **Col·lecció** [Garba \(Edhasa \(Firma\)\) ; 2](#)

# MARC record from Biblioteca Nacional

...

041 cat|alat|hlat 

080 0 80.1-14"-00"

100 10 [Catulo, Cayo Valerio](#) 

243 10 Obras|lCatalán-Latín

245 1



[Poemes|h\[Texto impreso\]|cCatul ; traducció \[del  
latin\] i notes de Josep Ignasi Ciruelo & Jaume Juan](#)

250 Edició bilingüe íntegra

260 0 Barcelona|bEdhasa|c1982

300 XVIII, 254 p.|c16 cm

490 [La Garba|v2](#)

....

**Translators??**

# Record at The European Library

OpenSearch: ...advanced((CONTRIBUTOR,Ciruelo)AND(YEAR...

## Poemes

ED. BILINGÜE ÍNTEGRA

### Harvard-Style Citation

Catullus, G. V., 1982. Poemes: ed. bilingüe íntegra. Barcelona:  
Edhasa

### Creator

Gaius Valerius. Catullus

### Contributor

Josep I. Ciruelo

### Audience

Unknown

### Edition

1. ed.

### ISBN

8435003744

### Published

Barcelona: Edhasa, 1982

### Resource type

Text

### Extent

XVIII, 254 S.

# Problems with metadata

- When harvesting some data is lost. Moving from fine grained descriptions (MARC) to DC.
- SRU: supported indexes do not include MARC. The cost of having big indexes, needs to be considered???. (LOC does not support MARC, BNE does not support dc.contributor).
- Researchers look for ‘specialized corpora’, so metadata needs to be addressed.

# UPF 'federated' content search

- Four different corpora
- Process:
  1. SRU for metadata search.
  2. Temporal sub-corpus (indexing)
  3. CQP for content search & query

- Hard to implement for big subcorpus but:
- Better for queries implying **sort**, **count** or **group-by** (they are part of the CQP query)
- No aggregation is required at this level
  
- FS implies massive data and massive results.
- Researchers cannot consume massive concordances. Massive data require statistical analysis methods to manage them (group-by, sort, count...)