

CRMS South America

A Study of Argentine monographs in the Benson Latin American Collection, University of Texas at Austin

Carlos Ovalle, Caron Garstka, Georgia Harper

I recommend that this presentation include a CC*BY license (or other CC license on which the authors agree).

Background

- University of Texas at Austin (UT) became a Google Books library partner in 2007
- Google scanned materials from the Benson Latin American Collection
- UT gave up access to scans of in-copyright works in 2009 ('settlement amendments')
- Project completed 2010; catalog maintenance continues

CRMS Project Goals

- Test University of Michigan (UM) CRMS on Latin American works
- Build a framework for risk analysis for non-public domain materials
- Provide a basis for UT Libraries to determine value of assessing public domain status
- Persuade Google to provide access to in-copyright scans for purpose of PD assessment only (ie, amend the UT contract)

Getting Started

- UM supplied informational resources and the CRMS model
- Personnel at UM and UT advised and assisted
- Researchers selected Argentine monographs from the list of Google scans
- Researchers developed and tested protocols and a form for data collection modeled on CRMS (because UM did not have UT catalog data)

Benson Latin American Collection

- ~1.2 million items in collection
- ~500,000 scanned by Google
- 18,918 monographs from Argentina

Argentina

- Relatively large sample size in collection
- Relatively simple copyright law

Assumptions

- Registration required, but ignored for our purposes because registration can occur any time prior to copyright expiration
- Assumed GATT applies
- Skipped compilations and translations

The consequence of failure to register is that others can basically do whatever they want with your work. If it is later registered, there is nothing the copyright holder can do about things that would have been infringing but for the failure to register.

This is of little practical value to us because of the technical impossibility at this time (we don't have access to our scans and even if we did, we would have to indemnify Google for its damages if it were sued, so we are quite unlikely to do this). But it is interesting to note. See Section 63, and the parallel reference for translations that are not registered (contracts not registered), Section 23.

Translations: have to be authorized while author is alive and for 10 years after death, but after that, anyone can do a translation without an agreement with the heirs (but it is implied that you have to try), and fees will be fixed by arbitration.

Time Involved

- Total project time: 5 months
- Library staff time: Creating list ~4 months!!
- Library staff time: Pulling books - 1 week
- Pulling: 10 books/hour *if books still on site*
- Entering catalog data: 16-20 hours
- Determining death dates: 8-16 hours

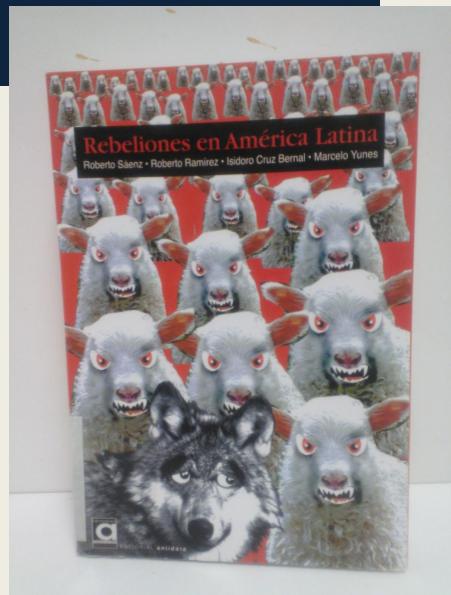
A Sample of the Record

Really helpful, isn't it...

	"RECORD #(BIBLIO)"	"BCODE3"	"BIB_LVL"	"CALL #(ITEM)"
"001"	"b1553859x"	"_."	"m"	"910.6 P191P"
"6382690"	"b15538606"	"_."	"m"	"910.6 P191P"
"7598591"	"b15540017"	"_."	"m"	"910.6 P191P" "910.6 P191P"
"5360567"	"b15540078"	"_."	"m"	"910.6 P191P" "910.6 P191P"
"7413154"	"b15540082"	"_."	"m"	"910.6 P191P" "910.6 P191P"
"16418338"	"b15540182"	"_."	"m"	"910.6 P191P" "910.6 P191P"
"16418948"	"b15540194"	"_."	"m"	"910.6 P191P" "910.6 P191P"
"12392693"	"b15543997"	"_."	"m"	"9123 D57 1984"
"643429"	"b15603192"	"_."	"m"	"QE 231 S47 1969" "QE 231 S47 1969"
"768714"	"b15616873"	"_."	"m"	"F 1413 H3464"
"877827"	"b15631394"	"_."	"m"	"BR 665 T65"
"108627"	"b15634414"	"_."	"m"	"1408.3 B967 1918"
"1086104"	"b15658324"	"_."	"m"	"F 1086.3 B668"
"1147166"	"b15666682"	"_."	"m"	"982 SEL5"
"1191260"	"b15672359"	"_."	"m"	"-Q HG 5312 A73 1971" "-Q HG 5312 A73 1971"
"1191296"	"b15672360"	"_."	"m"	"PS 111 I5 T918"
"1251507"	"b15679536"	"_."	"m"	"LA 548 L317"
"1699475"	"b15720524"	"_."	"m"	"HT 1123 C6 A87" "HV 5840 E2 P58 1988" "HT 1123
c6 A87"				
"1802953"	"b15728055"	"_."	"m"	"19 C67 1974"
"1904657"	"b15736180"	"_."	"m"	"JX 6570 A7 G372"
"1950557"	"b15740742"	"_."	"m"	"3316 V38"
"2032300"	"b15750346"	"_."	"m"	"JL 203 L554"
"203243"	"b15751686"	"_."	"m"	"HT 153 B6534" "HT 153 B6534"
"2052339"	"b15752466"	"_."	"m"	"Q 127 L38 F477" "Q 127 L38 F477"
"2065225"	"b15753438"	"_."	"m"	"JL 2011 L664 1975"
"2076517"	"b15755125"	"_."	"m"	"F 2235.4 A753"
"2118100"	"b15759398"	"_."	"m"	"JC 328.6 G883"
"2131063"	"b15761228"	"_."	"m"	"JL 2015 1974 P47"
"2139764"	"b15762300"	"_."	"m"	"JC 71 P6 E385"
"2153761"	"b15763572"	"_."	"m"	"F 1408.3 D518" "F 1408.3 D518"
"2162708"	"b15764199"	"_."	"m"	"QH 380 L38"

Some covers were interesting

Others, not so much...



Preliminary Results

Codes independently reviewed 100 books

- 64% had inserts
- 3% were translations
- 16% were compilations with many authors

Reliability

- One date difference between coders
- One author with 2 death dates
- Catalog was at least 99% accurate *

* Catalog data from older books often said, “et al” for multiple author entries because of the need to fit the entry on a card catalog. But there is only one entry where two authors were mentioned and the book itself gives no indication whatsoever of the second author.

Preliminary Results

Works definitely in the public domain

- 9 Argentina
- 8 US
- 4 Argentina & US

Reasons for discrepancies

- US uses date of publication for pre-1978 works;
Argentina uses author death date
- Shorter terms for government works, photographs,
anonymous works

Anonymous works (of people, or institutions (ie, works for hire with no identified person as author): 50 years

Photographs: 20 years

Government works, if they are anonymous, 50 years; with an identified author, life of the author plus 70 years.

Preliminary Results

Able to determine copyright expiration dates

- 31 Argentina
- 46 US
- 31 for both Argentina and US
 - Generally because pre-1978 US expiration is based on publication date while Argentina uses author death date

Preliminary Results

Death date specifics

- 39 items with sufficient information to conclude that authors are definitely or *probably* living
- 37 items for which we found definitive death dates
- 15 items where we found nothing definitive to confirm whether author is alive or dead
- 3 items were government or other entity-authored works without individual personal attribution

Author Death Date Information

Probably living = we found a recent (since 2000) reference online to the author's being active (publishing, presenting, teaching).

We had a publication date for every item. (100%)

We had only one item from the 1800s (1887).

The rest ranged from 1906 to 2005, with the great majority published after 1940 (88%).

This date range is probably due to characteristics of the collection combined with what Google scanned. The original collection came about in the 1921s, and was the Garcia collection (which is why we had that "G" letter in front of a lot of call numbers, Caron). At some oddly-difficult-to-determine point it became the Latin American Collection. Nettie Lee Benson became involved with it in 1942, and probably was responsible for a good deal of acquisition and their collection development policies. Google did not scan things that were in rare books, things that had margins that were too narrow, fragile material, and a few other criteria. Although it changed over time, it's almost certain that items prior to 1900 (for serials, and for monographs possibly later) were automatically considered rare books at the time Google scanned so and so Google didn't get a chance to scan them at all.

Things we found helpful

- Access to library staff who were involved in the Google Books project
- Access to Benson collection staff
- Access to a cataloger for general information
- Rudimentary knowledge of Spanish
- Google translate or other translation sources

Helpful Sites for Author Death Dates

- UT Catalog
- Biblioteca del Congreso (authoritative)
- Wikipedia
- Library of Congress Name Authorities
- Google Search
- Social Media (LinkedIn, Facebook)
- University Websites
- Newspaper articles
- Biografías y Vidas, Minibiografías
- Tango site...

Todo Tango



Some Challenges

- Spanish (comprehension;
accented characters affect search)
- Legal (GATT, translations, government works)
- Lost or destroyed records
- Staff turnover
- ILS migration
- Catalog quirks (et al. used to fit multiple authors info on
physical catalog cards)
- Collection changes over time
(some monographs moved to Library Storage Facility)

Card Catalog



Risk Analysis for non-PD works

What factors might affect decisions regarding what to do with non-PD works

- Author status (known/anonymous, person/institution)
- Publication date
- Inserts/other Contributors
- Edition, translation, country of publication, volume, etc.)
- Partial PD - not PD in every country

translation (whether or not authorized is not always clear); if his translation contract is not registered with the National Intellectual Property Register, within one year of the publication of the translated work, the author of the translation has no rights (to enforce it, I ASSUME). BUT he can recover his rights at any time by registering, and then he could enforce them, but he can't object to other translations made during the time his contract was not registered with the NIPR.

Future Work

- Continue study: Evaluate ~400 works for better sample data
- Ask Google for access to scans - the only viable basis for going forward
- Present findings to library administration for decision about long-term (including need for scans to conduct PD analysis)
- Create local CRMS, or integrate into UM interface
- Further develop risk assessment framework for decisions about limited access to non-PD works