

# Doing the LibraryThing™ in an Academic Library Catalog

Christine DeZelar-Tiedman

Archives and Special Collections Catalog Librarian, University of Minnesota Libraries, Minneapolis, Minnesota, USA

## ABSTRACT

### TITLE:

Doing the LibraryThing™ in an Academic Library Catalog

### BACKGROUND:

Many libraries and other cultural institutions are incorporating Web 2.0 features and enhanced metadata into their catalogs. These value-added elements include book jacket images, reviews, and user-generated tags. One such site that libraries are looking at as a model is LibraryThing. LibraryThing is a social networking site that allows users to “catalog” their own book collections. Members can add tags and reviews to records for books, as well as engage in online discussions about them. In addition to its service for individuals, LibraryThing offers a fee-based service to libraries, where institutions can add LibraryThing tags, recommendations, and other features to their online catalog records.

### OBJECTIVE:

To determine the potential value of importing user-contributed data from LibraryThing into a large academic library catalog, by analyzing the overlap between the two databases, and the quality and quantity of metadata available from each source.

### METHODS:

A random sample of 383 titles from the University of Minnesota Libraries catalog (MNCAT) was searched in LibraryThing. The characteristics of matched and non-matched records were examined. For this project, metadata analysis focused on access via subject terms. Comparisons were made between the user-supplied tags in LibraryThing (social tags) and the Library of Congress subject headings (LCSH) and Medical Subject Headings (MeSH) in the library catalog records (controlled vocabulary system).

### RESULTS:

Of a random sample of 383 titles from MNCAT, 80 works, or 21 percent of the sample, had corresponding records available in LibraryThing. Recent publications in English in the book format were strongly favored. There were a total of 227 LCSH/MeSH terms applied to the 80 catalog records, an average of 2.84 per record. In LibraryThing, 498 tags were applied to the same 80 titles, an average of 8.73 per title. Only one record had no subject metadata in either source. Popular user tags closely correlated with controlled terms roughly 31 percent of the time.

### CONCLUSIONS:

While overall LibraryThing is not a strong source of enhanced metadata for large academic libraries, for recent, popular materials the user-contributed metadata enhances and complements that which is found in a traditional catalog record. This is particularly true of literary works, which have not traditionally received subject access in library catalogs.

## RECORD MATCHING

Title: A practical view of the prevailing religious system of professed Christians ... contrasted with real Christianity / William Wilberforce.  
Matched, though MNCAT record is for microform.

Title: Islandfischer / Pierre Loti  
MNCAT record is for a German translation of Pêcheur d'Islande. Matched to LibraryThing record for original French title.

### Comparison of subject metadata

- Formal
  - LCSH/MeSH—controlled vocabulary consisting of a main term alone, or with a combination of sub-headings identifying further topical, chronological, geographic, form or genre subdivisions.
  - LibraryThing tags—Single free-text terms contributed by users.
  - In order to make a meaningful comparison, LCSH/MeSH terms were “faceted”, i.e. broken down into their component parts and counted separately. Each unique facet was counted only once per record.
- Assessment of terms
  - The most popular tag from LibraryThing was compared with the LCSH/MeSH terms to assess the correlation between terms assigned by professionals vs. end users.
  - LibraryThing tags of a personal nature (i.e. irrelevant to other users) were identified and eliminated.
  - Subject metadata for literary works were compared
  - No attempt to evaluate accuracy from either source

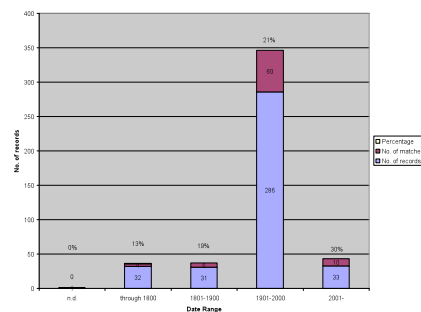
## RESULTS

### Match rate

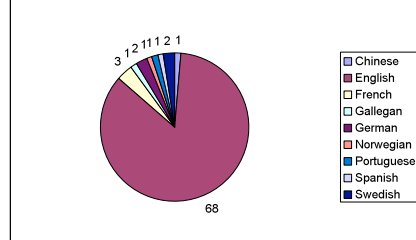
- 80 records, or 21 percent of the MNCAT sample, had a matching work in LibraryThing.
- Hit rates were highest for 20<sup>th</sup> century (21 percent) and 21<sup>st</sup> century (30 percent) materials.
- All of the matching titles were for books (monographs), though some of the MNCAT holdings were microforms or e-books.
- Of the 80 matching records, 68 (85 percent) were in English. The 12 remaining matches were for titles in Chinese, French, Gallegan, German, Norwegian, Portuguese, Spanish, and Swedish.
- 22 percent of English language materials in the sample were found in LibraryThing, as compared to 15 percent in all other languages.

## RESULTS

Record Matches by Date



Matches by language



### Subject Metadata Comparison

- Quantity of subject descriptors
  - LCSH/MeSH: 227 terms, an average of 2.84 per record
  - LibraryThing: 698 terms, an average of 8.73 per record
- Overlap and quality
  - Though 15 MNCAT records and 12 LibraryThing records had no subject descriptors, only one record had no subject metadata from either source
  - For 25 titles, the “most popular” LibraryThing tag roughly matched with a LCSH or MeSH heading for the same record. 27 titles did not have a corresponding close match. The 28 remaining titles lacked subject terms in one or both sources and so could not be compared.
  - When “personal” tags are eliminated, LibraryThing has 584 tags, or an average of 7.3 per record. However, some personal tags might be useful to some users.

## RESULTS

### Literary works

- 22 records in sample were for literary works
- 28 LCSH terms were provided, or 0.79 per record
- 293 user tags appeared on the LibraryThing records (13.32 per record)
- Eliminating personal tags, 230 LibraryThing tags are provided (10.45 per record)
- 14 MNCAT records for literature had no subject headings, vs. one in LibraryThing (3 if personal tags are eliminated)
- Due to the small sample size, further research is needed to draw strong conclusions

### Examples of “Personal” LibraryThing Tags

lebook	Hc/dj/1/1
3 <sup>rd</sup> shelf up	Lovely
8xx	Mer
Batch 14	Master bedroom
Beth	Owens
Borrow	Read
Bulk-m	Tbr
Carels	Unique
Ds	Unread
Daily_show	Useful for b00k
Favorites	Vf
First two sections	Wanted
Funny	Weed

## CONCLUSIONS

While overall LibraryThing is not a strong source of enhanced metadata for large academic libraries, for recent, popular materials and literary works the user-contributed metadata enhances and complements that which is found in a traditional catalog record. The following observations also need further study:

- User tags are often more specific or less specific than controlled-vocabulary terms. While some precision is lost, different levels of hierarchy may be appropriate for different levels of users.
- Tags incorporate the same or similar concepts in a variety of ways, using variant spellings, synonyms, singular and plural, and different parts of speech. This makes exact comparisons with controlled terms challenging.
- Some “personal” terms might actually be useful to a wider audience. Readers may gravitate toward a book that many have tagged as a “favorite” or may want to read something “funny” or discover which books were discussed on a particular television program.

For additional information please contact:

Christine DeZelar-Tiedman  
Technical Services Department  
University of Minnesota Libraries  
dez002@umn.edu

## OBJECTIVE

To determine the potential value of importing user-contributed data from LibraryThing into a large academic library catalog, by analyzing the overlap between the two databases, and the quality and quantity of metadata available from each source.

## METHODS

### Determining overlap between library catalog and LibraryThing

- Relative database size (as of 1 February 2008)
  - MNCAT: 5,200,000 records (n.b. some records are suppressed or otherwise invalid)
  - LibraryThing: 23,000,000 records

### Random sampling

A random integer generating program was used, to create a list of non-repeating integers from 1 to 5200000. Our Aleph500™ catalog system uses a sequential record numbering system beginning with 1. The record corresponding to each random number was called up in MNCAT, and the first 383 valid records were searched in LibraryThing for a matching record.

### Matching criteria

LibraryThing employs a “work” concept similar to that defined in Functional Requirements of Bibliographic Records (FRBR). That is, there is one LibraryThing record for all editions of an intellectual work. Thus, a LibraryThing record was considered a match if the same work was represented, regardless of the edition or format held by the library.