Providing access to literary works remains a challenge for catalogers and metadata librarians, notwithstanding modern advances such as the introduction of the Guidelines on Subject Access to Individual Works of Fiction, Drama etc. and the Library of Congress Genre/Form Terms. This study explores how harnessing the social cataloging of fiction and other belles-lettres might help meet this challenge. Samples of records from the catalogs of a university and a public library were compared with their equivalents in the LibraryThing (LT) platform, using a similar study reported in this journal ten years prior as a baseline. Most of the library titles were found in LT, and most were linked to tags that still offered additional access points of considerable value beyond the subject and genre headings included in the library records. However, the number of relevant and useful tags attached to each title varied considerably, as indeed did the quantity and quality of the headings. The authors analyze how the tags complemented the headings and identify genre, setting, theme, characters, and authorial attributes as key elements of description for social catalogers of literary works.

Social cataloging sites such as LibraryThing (LT) and Goodreads have given the reading public the opportunity not merely to use bibliographic records, but also to create them. In contrast to how library cataloging is sometimes viewed (as a rather arcane exercise), these sites have become remarkably popular. LT touts itself as a platform for “a community of 2,550,000 book lovers,” while Goodreads claimed around 90 million members as of July 2019.1 It would appear that any difficulties people might encounter with library bibliographic records do not dampen their enthusiasm for sharing details about their own collections. In some cases, this enthusiasm might not extend to much more than using existing records, which are typically based on library bibliographic records. However, in other cases, users of these sites are happy to also contribute their own cataloging, adding tags, reviews, and other elements to become social catalogers.

The contribution of these social catalogers makes to the cause of access provision started to be investigated not long after these sites were first launched, in the mid to late 2000s, as outlined in the literature review below. While these investigations have led to the general view of social cataloging complementing the work of professional metadata librarians, exactly how and to what extent the former adds value, potentially, to the latter depends on the context...
of particular systems, sources and materials. This context remains under-researched.

Belles-lettres is a category of material of particular interest in relation to social cataloging. Not only is it a category that is well represented in sites such as LT, it is also one that has been less well covered by traditional cataloging practices. However, few studies of the social cataloging of fiction and other literature have been published. This paper aims to help address this gap, presenting research that builds on a study reported about ten years ago in this journal, by DeZelar-Tiedman. 2

The new research questions whether the added value of LT tags for literature that DeZelar-Tiedman identified has increased or decreased in the intervening decade, and why. The authors are mindful of how LT and similar sites have grown their user base and coverage during this period, but also of the important developments in the library cataloging of belles-lettres in the 2010s, including the application of the Library of Congress Genre/Form Terms (LCGFT), a new list of headings covering literary works, among other materials. The current study also extends the earlier study by comparing tags for works owned by both a public and a university library, and by conducting additional analysis on the new samples. While libraries continue to grapple with the best ways to leverage social cataloging, library professionals need to ask ourselves whether this has become less (or more) of a pressing issue, at least for particular types of resources.

**Literature Review**

**Social Tagging and Social Cataloging Studies**

Interest in user-generated metadata predates the social cataloging sites that began to appear in the mid- to late 2000s, with calls for “democratic indexing” to address the limitations of library cataloging first made more than a decade earlier. Most notably, Hiderley and Rafferty argued that the tagging of artistic and literary works by the public at large could be particularly effective if structured according to the various levels of meaning of works of the imagination. 3 However, it was the advent of “Web 2.0” and an online environment that readily accommodated user tagging that established social tagging as a major area of research in library and information science (LIS). Scholars such as Mai, and Pando and Almeida, championed social tagging’s postmodernist approach to knowledge organization, affording multiple viewpoints over the singular perspective of an intermediary or expert. 4

Social tagging research has been summarized by Rafferty, who notes that “social tagging generally means the practice whereby internet users generate keywords to describe, categorize, or comment on digital content.” 5 However, tags can also be added to records of physical objects, just as headings are added to records for these resources created in library catalogs; this practice has become known as social cataloging, a particular subset of social tagging. 6 Other categories of social tagging have also been identified, such as the those that relate to how different platforms and systems allow for different degrees of participation. Some platforms, including LT, allow all site users, or at least all subscribers, to contribute tags for a given resource, whereas others enable only the contributor of the resource, or particular categories of user, to do so. 7

One strand of social tagging research has explored the different motivations of taggers, and the kinds of tags that result from them. Some tags may be less helpful in facilitating access to the wider community, and are more of a “personal” nature, representing an idiosyncratic relationship between tagger and resource of little relevance for the user population in general. 8 The extent to which users are prepared to tag may well vary according to the function and purpose of the platform being used.

In cases where tags can be generated in significant quantities and where many of these tags are added to provide access for users at large, their value has been generally regarded as complementary to that of any controlled indexing added by information professionals such as library catalogers. 9 Strengths identified include tags’ flexibility and currency, plus their broader accommodation of multiple viewpoints and terminologies. They tend to provide a greater level of “recall,” but a lower level of “precision,” to use the classical measures of indexing quality. 10 Social tagging is usually a much cheaper option than professional indexing; indeed, in many contemporary file sharing environments, it may be the only option that is sufficiently scalable. 11 However, despite the apparent inclusive nature of social tagging, studies have shown how a relatively small number of “supertaggers” tend to produce the lion’s share of many “folksonomies” (i.e., the social tagging aggregations). 12 Regardless, no indexing, whether produced by a large number of people or by a single person, is devoid of ideology and bias. 13

Since the 2000s, the complementarity of library cataloging and social tagging has been explored in various studies, many of which have focused on the nature and value of social cataloging. A majority have made use of the publicly accessible, and popular, LT site. Typically, social cataloging tags are compared with headings created by library supplied cataloging for the same resources. In an early study, Heymann and Garcia-Molina found that many, though by no means all, of the tags in the LT and Goodreads sites were of value from an access perspective. 14 They also found that tags and library-assigned subject headings were assigned in similar frequency ratios, suggesting a similar “depth” in
the social and professional approaches. Rolla compared the tags from LT and the Library of Congress Subject Headings (LCSH) on OCLC WorldCat records for a small sample of common materials, and found both similarities and differences. There were many more tags than there were headings, even in this early period of social cataloging, and many tags that did not correspond to the terms in the headings; some of the tags were personal in nature, while others were more descriptive. In contrast, for each item there were some concepts, at a broad level, that could be identified in both the tags and headings. A breakdown of the less useful and more “personal” tags to be found in LT was provided by Lawson, and includes reading status, date, gift suggestion and location of the copy.20

Adler focused on the LT tags used in “transgender books,” showing how they often differed in both language and concept from the headings assigned by libraries. Meanwhile, Thomas, Caudle, and Schmitz compared the tags and headings for ten popular books, employing a taxonomy that covered certain semantic relationships, such as broader and narrower terms. A similar study of tags for a wider range of academic library materials was conducted by Voorbij, using a sample of 160 records, while DeZelar-Tiedman compared the LT tags and subject headings specifically for sequences of literary works in a university library collection.19

DeZelar-Tiedman’s study focused on how the retrieval value that LT tags might add to the headings already provided in her university library’s catalog. Samples of records for literary works by twentieth- and twenty-first-century American and British authors, which had been classified as such, were collected from the catalog; their headings were then compared with the tags assigned to matching works in LT, or at least with the works’ thirty most frequently assigned tags, where applicable. DeZelar-Tiedman found that 43.0 percent of the sampled works contained LT tags, but no LCSH, while a further 33.8 percent contained both LCSH and LT tags. Among a sub-sample of fifty works linked to both LCSH and LT tags, there were numerous instances of the complementary nature of the headings and tags, providing “a more complete view of the nature and thematic elements of the work than either sources does alone.”20

Iyer and Bungo based their analysis on a sample of books in complementary and alternative medicine, and classified both the tags that related and that did not relate to the headings into various broad categories, such as “time period” and “locations.”21 A more statistical approach was taken by Lu, Park, and Hu, who analyzed the overlap between LT tags and LCSH using the Jaccard index, counting the terms in common divided by the total number of terms, based on lists of the most frequently occurring tags and headings. The very limited degree of overlap, and the large dataset, suggested considerable potential for LT tags to complement LCSH. In a more practical study, Pirmann conducted a usability analysis of a library catalog augmented with social tagging. She found that the participants made good use of both headings and tags, for different purposes, confirming their complementary nature.

Šauperl conducted an extensive study of different ways that novels were described by publishers, librarians, literary theorists, and readers. In relation to the last category, she examined the tags and reviews added by users of LT and the Amazon bookstore for twenty well-known novels. She identified that the LT tags covered a number of elements for all the novels: literary character, genre, topic, author, position in literary history, and place. There was a strong correlation between the LT and Amazon tags’ coverage of elements, while a few elements were prevalent across the descriptions produced by publishers, librarians, literary theorists and readers alike: story, information about the author, genre, personal experience with reading the novel, and evaluation.

Pecoskie, Spitteri, and Tarulli compared the tags and headings in Canadian public library catalogs for award-winning fiction. In this case, the tags were entered directly into the catalog by library users, rather than in a social cataloging platform. The analysis revealed differences in the distributions of tags and headings across the typology the authors constructed, with proportionately more tags representing awards, tone, and topic, and with more headings representing genre, location, and period. The authors also compared their results with the elements included in models of fiction description previously put forward for readers’ advisory services, as listed below, finding some discrepancies.

<table>
<thead>
<tr>
<th>Award/recognition</th>
<th>Genre</th>
<th>Setting</th>
<th>Real events</th>
<th>Factual information</th>
<th>Pacing</th>
<th>Paratext</th>
<th>Specific characters</th>
<th>Characters’ occupations</th>
<th>Time</th>
<th>Plot development</th>
<th>Ending</th>
<th>Readability</th>
<th>Advice to readers</th>
<th>Emotional experience</th>
<th>Subject</th>
<th>Characters’ relationships</th>
<th>Intended audience</th>
<th>Library influences</th>
<th>Size/length of book</th>
</tr>
</thead>
</table>
The use of social tagging in library catalogs has been promoted by these authors in other publications, including Tarulli’s *The Library Catalogue as Social Space*.26

There have been fewer studies of social cataloging in more recent times. Vaidya and Harinadrayana performed an analysis similar to that of Lu, Park, and Hu, that focused on LIS materials, and found a similarly low level of overlap, as did Samanta and Rath in their study of LT tags in the field of economics.27 Michael and Han examined the tagging in an academic library catalog over a seven-year period, finding uneven coverage, with some tags of promise, and others of lesser utility.28 Hider searched the LT site for tags representing various fiction genres listed on Wikipedia that were not on the LCGFT list, and compared their presence in LT with those representing a sample of genres that were included in LCGFT, finding the former to be more prevalent than the latter.29

**Library Cataloging of Belles-Lettres**

One category of material that has been less well served by library supplied cataloging are works of the imagination, including literary works or belles-lettres. Indeed, until recently, the cataloging of fiction and other belles-lettres tended to be quite minimal. Literary works were classified and indexed with reference to a few broad facets, such as form, language, nationality, and period, but subjects and genres were not addressed, partly because of the difficulties catalogers might face in determining them. This approach, however, severely limited readers’ access to their libraries’ literature collections, leading to a number of initiatives over the past thirty years to address this deficiency.30 Of particular note is the American Library Association’s development of its *Guidelines on Subject Access to Individual Works of Fiction, Drama etc.* (GSAFD), which provided a framework for subject-related access points to be created by catalogers of literary materials, and then the development of the LCGFT, begun in 2007, which now covers a wide range of materials, including literature, and others such as music and film.31

These developments have led to changes in cataloging practices and fuller levels of bibliographic records for fiction and other literary works, including the analysis and indexing of literary subjects and genres. However, the extent to which practices have changed across the library community as a whole is unclear. Likewise, it is not clear that the new practices, where they do occur, align with the findings from those studies that have explored how readers seek out literary works. While findings from these studies have highlighted the diversity of methods and strategies employed by readers to this end, they also identified several core elements for search systems to cover. For instance, Beghtol proposed that the facets of characters, events, spaces and times are “fundamental data categories for fiction,” while Ranta noted the need to cover both denotative and connotative elements, and Pejtersen and Austin found that public library users sought fiction according to the four basic elements of subject matter, frame, author’s intention and accessibility.32

Many of the empirical studies of how readers sought out fiction and other literary materials were based on interviews and surveys. However, the provision of access to fiction outside of librarianship has also been analyzed to help inform the enhancement of access to literary collections within libraries. Adkins and Bossaller compared the access points to fiction provided in online bookstores, readers’ advisory databases and library catalogs, finding that together the sources covered a wide range of elements, with the different platforms offering complementary means of access to a significant extent.33 Elements that were frequently identified across the platforms included emotional experience, explicit content, factual information, specific characters, characters’ occupations, characters’ relationships, setting, time, plot development, pacing, and subjects. Recently, Hider and Spiller mapped the fiction genres used in online bookstores and Wikipedia to those of LCGFT, revealing many discrepancies between the commercial and library vocabularies, and also amongst the bookstores, some of which appeared to be based on geography.34

**Method**

This paper reports on a study that replicated and extended the study conducted by DeZelar-Tiedman.35 The focus was likewise on user tags employed in LT to describe and potentially enhance access to belles-lettres, but ten years later. Two new samples were collected: the first was derived from the same source of bibliographic records as those in the earlier study (the University of Minnesota Libraries catalog, MNCAT), and the second was based on the run of adult fiction arranged alphabetically by author in the nearby Madison (WI) Public Library.36 The samples were collected in October 2019 and March 2020, respectively.

Both samples were somewhat random and derived in similar fashion, and similar to how the sample in the earlier study was collected. For the belles-lettres in the university collection, the two call number sequences used by DeZelar-Tiedman, based on LCC, PR6001-6126 and PS3500-3626, were displayed in MNCAT. The sequences cover the modern works of English and American literary authors. As in DeZelar-Tiedman’s study, the 125th record listed was identified for the sample, though in this study, the records were counted in reverse from the end of the sequences. Again, as in the earlier study, “literary criticism” and “publications that collected or compiled works of literary authors that
were originally published separately" were excluded from the sample.37

A total of 330 records were included in the first sample, compared with the 444 records DeZelar-Tiedman used. Criticism and collected works were excluded from the sample after the counting in the current study, but prior in the case of the earlier study, which would have contributed to the smaller size of the new sample. The overall population is likely to have grown, of course, though some weeding might have occurred, while possible changes to the catalog's call number browse function might also have been a factor. Regardless, the sample was deemed sufficiently close in size to that of the earlier study to allow for comparison.

With the public library collection, its various forms of belles-lettres were scattered across different sequences, with adult fiction arranged separately. Given that fiction was the predominant literary form in both university and public library collections, the authors focused on this sequence and collected a sample in a similar way, by means of its online catalog. Thus, every 125th record was identified for the sample, until 400 records were identified, after which collected works were excluded. It should be noted that this sequence included small numbers of fiction originally written in languages other than English. A total of 346 records were included in the second sample.

The following elements in the records of both samples were recorded in Excel spreadsheets: title, author, year of publication, and the number of subject and related (MARC 6xx) headings, excluding foreign language headings, divided by subject vocabulary (LCSH, LCCFT, FAST, etc.). Each title was then searched in LT. (This social cataloging platform was chosen to facilitate comparison with the DeZelar-Tiedman study, although Goodreads appears to be the more popular platform nowadays, and could well be worth using as an alternative basis for future research in this area.) When matching works were found, the tags used for the first match listed (by “relevance”) were copied into the Excel spreadsheets with each of their frequencies (i.e., the number of times the tag had been assigned by different users to the work). All tags assigned for each work were counted and recorded, as tags that were assigned more than once (i.e., by multiple users) to each work. Links to the catalog record and to the LT record were also included on the spreadsheets.

Following the earlier study, sub-samples of fifty titles were created for more detailed analysis. While literary form was used to structure the sub-sample in the previous study, the titles for this study were randomly selected from all the titles with LT tags in each of the two new samples. This was because the public library sample was limited to fiction. The distribution of form across the university sub-sample reflected that of its parent sample fairly closely (mean percentage difference being 6.8 percent). For each title in the two sub-samples, its LT tags were compared to its headings in similar, though modified, fashion to that of the earlier study, in which each tag was run against all the recorded headings for the title. In this study, each tag was compared with all the headings’ subfields, as this was deemed a more equivalent unit of analysis than the whole heading. As with the earlier study, in cases of titles with more than thirty tags, only the thirty most frequently used tags were analyzed. Each tag was categorized using an expanded version of DeZelar-Tiedman’s scheme, which had consisted of the five categories of Exact Match (M), Partial Match (PM), No Match: Specificity (NS), No Match: Vocabulary (NV), and No Match: New (NN), as defined below. The additional categories were used for those tags that fell outside of DeZelar-Tiedman’s scheme, such as those that did not associate with subjects. These other categories, namely Multiple Subjects (MS), Mixed (MX), Not Subject: Personal and Bibliographic Description (NSM), Not Subject: Personal (NSP), and Not Subject: Bibliographic Description (NSS), are likewise defined below. In cases of multiple applicability, the tag was recorded in the highest-listed category. Tags that described a format of the work not represented by the bibliographic record were set aside.

Multiple Subjects (MS) = the tag consists of multiple terms pertaining to one or more of the following five categories, but which do not articulate as a single compound concept.

Exact Match (M) = the tag matches exactly (except for capitalization) a subfield, and only that subfield, of a recorded heading. The subfield could be the first subfield of a heading.

Partial Match (PM) = the tag matches all the words of a subfield, and only that subfield, but the form of at least one of the words varies (in terms of spelling, hyphenation, plural/singular, verb vs. noun, etc.).

No Match: Specificity (NS) = the tag’s meaning relates to the concept represented by one or more subfields in the recoded headings, but not at the same level, i.e., is either more general or more specific (or both).

No Match: Vocabulary (NV) = the tag’s meaning is synonymous or near-synonymous with a subfield of a recorded heading, but uses one or more different words.

No Match: New (NN) = the tag represents a concept not covered by or hierarchically related to any
of the concepts in any of the headings and their subfields.

Mixed (MX) = the tag includes elements pertaining to one or more category above and one or more category below.

Not Subject: Personal and Bibliographic Description (NSM) = the tag covers both of the categories below.

Not Subject: Personal (NSP) = the tag serves a personal function for the tagger.

Not Subject: Bibliographic Description (NSB) = the tag is potentially covered in other parts of the bibliographic record and not by LCSH or LCGFT.

Not Subject: Space (NSS) = the tag consists of a space only.

Not Determined (ND) = the category for the tag could not be confidently assigned, i.e., its meaning was unclear.

The two authors categorized the tags for the first six titles in parallel, comparing and discussing their classifications after each title. After the sixth title, agreement reached a level of 97 percent, and the second author proceeded to categorize the remaining tags on her own.

To gauge the usefulness of the non-matching tags that pertained to subject (PM, NS, NV, and NN), the relevant tags were then evaluated according to the following three-point scale: “adds considerable value to the headings,” “adds some value to the headings,” or “adds little or no value to the headings.” Clearly this scale allowed for considerable variation in its interpretation and application, but the same author rated the tags across the two sub-samples, so that it could be used as a means of broad comparison. It also facilitated the identification of good examples of particularly useful tags, for the supplementation of library bibliographic headings, and of examples where their value in this respect was limited.

Modifying the last component of the earlier study, the nature of the non-matching, new-concept tags (NN) was analyzed and compared across the two sub-samples. Similar, though slightly different, categories were used for a first pass. Whereas DeZelar-Tiedman identified forms, genres, topical, geographic, chronological, and characters, mirroring LC structures, the authors decided to treat genres as part of an “abstract noun” category, given that genres were not always readily distinguishable from topics. In contrast, abstract and concrete nouns are more distinguishable, and the authors felt that this distinction might be of interest. Place name and personal name categories generally corresponded to the earlier study’s geographic and character categories. Additionally, an “affective” category was used to investigate the extent to which reader experience was explicitly indicated, outside of indications of genre. A “discipline/field” category was introduced to distinguish these tags from those of topic, form, and genre. Another departure from the earlier study was that the same sub-sample was used this time, whereas the tags analyzed in this component of the previous study were those linked to the works for which no LCSH had been assigned. Tags were recorded in the highest-listed applicable category.

Combination (M) = the tag consists of multiple terms that pertain to more than one of the categories below

Affective (A) = the tag indicates one or more emotions that the work elicits (often adjectively)

Discipline/field (D) = the tag indicates the discipline/field(s) of study to which the work belongs

Form (F) = the tag indicates the form(s) or genre(s) of the material

Subject: Common Abstract Noun (CA) = the tag contains a common abstract noun

Subject: Common Concrete Noun (CC) = the tag contains a common concrete noun

Subject: Place Name (PL) = the tag contains a place name (proper noun)

Subject: Person Name (PE) = the tag contains a person’s name (proper noun)

Subject: Other Proper Noun (PO) = the tag contains a proper noun that represents neither a place or person

Other (O) = the tag contains a concept not belonging to any of the above categories

Given the complexity of what “subject” means in the case of literary works, a second pass was conducted, in which the tags were coded inductively.

Findings

Table 1 shows the different genre distributions of the old and new MNCAT samples. The new sample included
considerably more material classified under “other,” proportionately. This may be partly due to a looser interpretation of the excluded categories (literary criticism and collected works). While the previous study indicated differences in the match rates of different genres, table 2 makes it clear that match rates have increased over the past decade in all genres, with the exception of short stories. Overall, coverage has increased from about 80 to 90 percent, reflecting how the social cataloging platform has increased its user base and coverage over the past decade. Of particular note is the 98.1 percent match rate for novels in the new sample. Even poetry, the genre with the lowest match rate, has almost three quarters of its instances covered on LT.

The sample of adult fiction from the Madison Public Library mostly consisted of novels, as might be expected, though 4.0 percent were short stories, and 1.7 percent “other.” This may have been one reason why its overall match rate was particularly high, as table 3 shows, although it also seems likely that the sort of material collected by the public library would be especially likely to show up on the LT platform due to greater popularity and accessibility (especially if it is stocked in public libraries). Its greater overall currency could also be a factor.

A breakdown of the presence of subject and related headings in the catalogs versus tags in LT is shown in table 4. Although current cataloging practices have increased the proportion of MNCAT records with headings, as table 4 shows, there are still about a third without headings, but with corresponding tags in LT, suggesting that there remains a strong case to explore the use of social tagging to enhance access to literary collections in academic libraries. With the Madison Public Library, all sampled records came with one or more headings, while the majority of the 98.3 percent matching titles in LT were tagged. Again, the high proportion of titles with tags could likely be attributed to the mainstream and contemporary nature of many of these titles, while the pervasiveness of the headings may be due to contractual agreements with the library’s suppliers.

Of course, the presence of headings or tags does not reveal anything about their number or quality. Table 5 provides a picture of typical quantities of headings and tags, with the median number of headings in the new sample of MNCAT records and the median number of tags in the matching LT titles broken down by genre. Whereas the MNCAT records typically include one or two headings, perhaps divided into two or three of four subfield elements, corresponding LT titles are typically linked to much larger numbers of tags. However, these numbers vary greatly across the sampled works, as shown in the breakdown. Novels are typically assigned many more tags than are poetry anthologies, no doubt partly because they are tagged by many more users on average. Similarly, even within these categories, quantities vary enormously. The full range for the 298 titles that match up with the new MNCAT sample is 0–3,303 tags. It should also be noted that the number of tags per record tends to drop substantially if tags assigned by only one user are discounted: the highest number of tags for the 298 titles is then 689, with the median dropping from 27 to 5.

Table 6 demonstrates that these quantities hide a very marked trend in library cataloging towards providing

---

**Table 1. Genres in MNCAT samples**

<table>
<thead>
<tr>
<th>Genre</th>
<th>DeZelar-Tiedman Study</th>
<th>New MNCAT Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Novel</td>
<td>244 55.0</td>
<td>154 46.7</td>
</tr>
<tr>
<td>Drama</td>
<td>96 21.6</td>
<td>27 8.2</td>
</tr>
<tr>
<td>Poetry</td>
<td>45 10.1</td>
<td>73 22.1</td>
</tr>
<tr>
<td>Short stories</td>
<td>38 8.6</td>
<td>24 7.3</td>
</tr>
<tr>
<td>Other</td>
<td>21 4.7</td>
<td>52 15.8</td>
</tr>
<tr>
<td>Total</td>
<td>444 100</td>
<td>330 100</td>
</tr>
</tbody>
</table>

**Table 2. Match rates for the MNCAT samples**

<table>
<thead>
<tr>
<th>Genre</th>
<th>DeZelar-Tiedman Study %</th>
<th>New Lit Study %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Novel</td>
<td>89.9</td>
<td>98.1</td>
</tr>
<tr>
<td>Drama</td>
<td>68.8</td>
<td>88.9</td>
</tr>
<tr>
<td>Poetry</td>
<td>68.9</td>
<td>74.0</td>
</tr>
<tr>
<td>Short stories</td>
<td>98.5</td>
<td>95.8</td>
</tr>
<tr>
<td>Other</td>
<td>81.0</td>
<td>88.5</td>
</tr>
<tr>
<td>Total</td>
<td>82.7</td>
<td>90.3</td>
</tr>
</tbody>
</table>

**Table 3. Overall match rates across the three samples**

<table>
<thead>
<tr>
<th>Sample</th>
<th>N</th>
<th>Matches on LT</th>
<th>Match %</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeZelar-Tiedman study</td>
<td>444</td>
<td>367</td>
<td>82.7</td>
</tr>
<tr>
<td>MNCAT 2019</td>
<td>330</td>
<td>298</td>
<td>90.3</td>
</tr>
<tr>
<td>Madison PL</td>
<td>346</td>
<td>340</td>
<td>98.3</td>
</tr>
</tbody>
</table>

**Table 4. Headings vs tags in the three samples**

<table>
<thead>
<tr>
<th></th>
<th>DeZelar-Tiedman Study (%)</th>
<th>New Lit Study (%)</th>
<th>Madison (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No headings, no LT tags</td>
<td>5.0</td>
<td>5.5</td>
<td>0.0</td>
</tr>
<tr>
<td>No headings, not in LT</td>
<td>13.1</td>
<td>4.8</td>
<td>0.0</td>
</tr>
<tr>
<td>Headings, no LT tags</td>
<td>1.0</td>
<td>2.7</td>
<td>2.6</td>
</tr>
<tr>
<td>Headings, not in LT</td>
<td>4.3</td>
<td>4.8</td>
<td>1.7</td>
</tr>
<tr>
<td>No headings, but LT tags</td>
<td>43.0</td>
<td>33.9</td>
<td>0.0</td>
</tr>
<tr>
<td>Headings and LT tags</td>
<td>33.8</td>
<td>48.2</td>
<td>95.7</td>
</tr>
</tbody>
</table>
greater access to literary works. If the new MNCAT sample is chronologically divided, the very sparse subject headings assigned to works published and generally acquired before the 1990s can be contrasted with a dramatic increase through the 2010s. Indexing for the subjects of fiction and other imaginary works has become much more common in many libraries, while the adoption of LCGFT has clearly also occurred in the case of the University of Minnesota, with 68 LCGFT headings in the new sample. Likewise, there are 442 LCGFT headings in the Madison Public Library sample (most of which would have been published in the 2010s). In addition to the greater numbers of LC headings, many of the more recent records included more non-LC subject headings, though some of these may overlap with the LC headings.

Whereas large proportions of literary titles in LT may be linked to more tags than their corresponding titles in library catalogs are linked to headings and subdivisions, the question of quality remains. Results from the analyses of the three sub-samples of fifty works provide an indication of the tags’ value. Table 7 provides the two new sub-samples’ breakdown across all the categories used in the revised and expanded scheme (see “Method” section). Significantly more tags, proportionately, are related to subject or genre in the MNCAT-based sub-sample: 70.8 percent compared with 56.6 percent, excluding the few tags that were “mixed.” Otherwise, the distributions are quite similar, with the largest numbers in the categories of “No Match: New,” “No Match: Specificity,” and “Not Subject: Personal.” These results indicate large numbers of tags with the potential to supplement subject and genre access to literary works, as was found in the earlier study.

The numbers in the categories used in the earlier study’s scheme were also compared across all three sub-samples. These are shown as percentages in table 8, with all the non-applicable tags discounted. There are similar percentages for the new concept category, i.e., about half. The percentages for the exact matches are also similar, but there are about three times as many specificity variants in the new sub-samples, and many more partial and vocabulary variants in the earlier sub-sample. Much of the variance could be due to the matching being done at the subfield level in the case of the new sub-samples. The distributions of the two new sub-samples are very similar. Overall, tables 7 and 8 suggest that tagging behavior, in terms of the nature of the tags, does not vary all that much across time and material, at least within the literary realm.

Given the similarity of the tag type distributions of two new sub-samples, broadly similar levels of value that the non-matching subject-related tags add to their respective catalog record headings might be expected. Table 9 does not confirm this for certain, bearing in mind the subjective nature of the rating scale used, but neither does it suggest otherwise. Across both samples, the percentages indicate that many of the non-matching tags, over half of which pertain to subject or genre, would add significant value for access purposes. Some of the specific ways in which these tags could do this are discussed at the end of this section.

The deductive coding for the final part of the analysis is summarized in table 10. The types of new concepts are distributed quite similarly across the two new sub-samples, except for forms and common concrete nouns. Given the focus on one particular form, i.e., fiction, in the Madison Public Library sample, the former exception is to be expected. The reason for the greater proportion of concrete nouns among the Madison tags is less clear, and invites further
investigation. The percentage of tags directly indicating emotions derived from the reading experience is small, while those for the other categories are broadly in line with the results of the earlier study. The “other” category comprised a mixture of tags, many of them adjectives, such as “contemporary,” “colorful,” “interwar,” and “light.” Clearly these terms related to a range of different aspects of the various works, precipitating the inductive analysis of the sub-samples, as described below.

The MNCAT sub-sample was grouped into new categories first, and then the Madison sub-sample, after which the categories were reconciled, resulting in those covered in table 11. Some tags could not be confidently included in any of the resulting categories, and were set aside; this could have affected the percentages a little, but they are reported here indicatively. The distributions of tags across the various categories were broadly similar for the two sub-samples.

The large proportion of tags that represent forms and genres not covered by the library headings confirms the initial coding. For both belles-lettres in general and fiction in particular, what the work is, as opposed to what it is about, is important for social catalogers, with their contributions adding to those forms and genres covered in the catalog records at different levels: some of the additional tags represented basic forms, such as “fiction” and “poetry,” some more specific genres, such as “thriller,” “suspense,” and “fantasy,” and others more specific still, indicating particular sub-genres or hybrid genres, such as “amateur detective” and “romantic suspense.”

The next two major categories of tag not covered by the library headings are for setting and theme. Again, while some headings for setting would have been included in some of the corresponding bibliographic records, there is clearly room for more from a social cataloging perspective. Regarding period, many of the more generic indications such as “20th century” and “1960s” have an LC equivalent. Other tags are less readily translatable, and show how social catalogers may be underestimating, however, the degree to which settings are thus more than incidental.

Some of these settings might not have been identified by the library cataloger as significant, which may be symptomatic of a relative disregard for setting, as an element that is less clearly related to subject. Library catalogers may be underestimating, however, the degree to which social catalogers read fiction for their settings and the degree to which settings are thus more than incidental.

The tags for “new” concepts that pertain to “theme” are of particular interest. They tend to be less obvious than some of the others, such as those for geographic place, and in this way may be of particular value. It may be easier to find novels set in Japan, for example, than to find novels that deal with “motherhood,” “grief,” or “aging.” Some of the tags, such as “ambiguous morality” and “second chances,” are also less than readily translatable into LCSH, though most are covered by the vocabulary, including those for some relatively obscure or ill-defined concepts, such as “Afrofuturism,” “betrayal,” and “hope.”

The next group of tags not covered by the library headings indicated various characters featured in the narratives. A few were specific characters, but a larger proportion were types of characters, such as “Native Americans” and “fathers.” Again, most of these could be covered by LCSH, but had not been in these cases. There were also a number of non-human characters (animals, ghosts, etc.).

The only other major category, representing more than 5 percent of the tags for “new” concepts, is a loose grouping around authorial attributes. Many of these tags indicated the author’s nationality and thus the literary “tradition” of which they were a part, in a very broad sense, such as “American literature” and “US poetry.” However, there were also tags that denoted other attributes, such as gender (e.g., “women writers”) and race (“author of color”). This aspect already features quite strongly in library cataloging.
in headings and classification, and in other elements of the bibliographic record, but perhaps not systematically enough.

Two other minor categories are included in table 11. Social catalogers indicated the intended audience for many of the works covered by the Madison sub-sample, as might have been expected, given the library’s inclusion of more materials for younger readers. There were also a few tags providing additional access to narrative style, but not so many, bearing in mind that this element is not given much attention in library cataloging. Examples that appeared potentially useful included “epistolary novels,” “dark,” and “first person.” Some of these terms are covered in LCGFT. A few other categories were also included in the second round of coding, but at even lower levels of frequency, including a category similar to the “affective” category used in the initial coding.

<table>
<thead>
<tr>
<th>Aspect</th>
<th>MNCAT</th>
<th>Madison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form/genre</td>
<td>39.0</td>
<td>33.7</td>
</tr>
<tr>
<td>Period</td>
<td>9.1</td>
<td>5.8</td>
</tr>
<tr>
<td>Place</td>
<td>5.1</td>
<td>6.1</td>
</tr>
<tr>
<td>Type of place</td>
<td>1.6</td>
<td>3.8</td>
</tr>
<tr>
<td>Occupational environment</td>
<td>0.5</td>
<td>5.5</td>
</tr>
<tr>
<td>Setting</td>
<td>15.8</td>
<td>21.2</td>
</tr>
<tr>
<td>Human condition</td>
<td>6.1</td>
<td>8.1</td>
</tr>
<tr>
<td>Human problem</td>
<td>6.4</td>
<td>4.9</td>
</tr>
<tr>
<td>Ideology</td>
<td>2.9</td>
<td>0.3</td>
</tr>
<tr>
<td>Theme</td>
<td>15.5</td>
<td>13.1</td>
</tr>
<tr>
<td>Specific character</td>
<td>1.9</td>
<td>2.0</td>
</tr>
<tr>
<td>Type of character</td>
<td>4.3</td>
<td>4.9</td>
</tr>
<tr>
<td>Non-human characters</td>
<td>1.9</td>
<td>3.2</td>
</tr>
<tr>
<td>Characters</td>
<td>8.0</td>
<td>10.2</td>
</tr>
<tr>
<td>Authorial attribute/tradition</td>
<td>13.3</td>
<td>6.7</td>
</tr>
<tr>
<td>Audience</td>
<td>0.3</td>
<td>4.7</td>
</tr>
<tr>
<td>Narrative style</td>
<td>2.4</td>
<td>2.0</td>
</tr>
</tbody>
</table>

How LT Tags Complement LC Headings

While some of the LT tags judged more useful in their complementarity covered areas that might be regarded as weaknesses in library cataloging’s treatment of literary works, other tags covered omissions in catalog records that might not have been expected. Such omissions may have been oversights or errors or the result of minimal cataloging practices. When the library cataloging was minimal, the LT tags often highlighted the seriousness of this detriment, covering concepts of great topic interest, such as “artificial intelligence,” or concepts about which it is hard to find literary works, such as “impersonation.”

It was noticeable that even many of the more recent works had not been assigned LCGFT, hence the large number of “new concept” tags pertaining to genre, including very mainstream genres. It is hoped that this will be less of a weakness on the part of library cataloging in the future, though it should be noted that a work can belong to a number of different genres, some of which might only be identified as such by a minority of readers, and that “proto-genres” may not have found their way into LCGFT. Indeed, there were a number of tags for which it was assumed were intended to represent genres, but that are not presently in LCGFT, such as “romantic suspense.”

Other concepts, particularly those representing themes, may remain less likely candidates for cataloging, due in part to the difficulty of identifying such concepts unless they are explicit in the secondary sources at hand (e.g., in the “blurb”). Indeed, not only might they not be identified, but even if they are identified, it can be difficult for the cataloger to assess their centrality to a given audience. Themes that emerged in the sub-sample, such as “family secrets,” could be a challenge in this respect. It should be noted that not all themes that might add value need to be abstract; indeed, more concrete examples, such as “rain,” might be of particular retrieval value because of their clarity of meaning.

In summary, there were tags for a wide range of concepts not covered by the library subject headings. Some of these covered gaps due to minimal cataloging practices; others were due to what may have been cataloger oversights or practical limitations to subject analysis; still others were possibly due to the narrowness of the cataloger’s singular viewpoint; others due to the constraints of the rules and policies for the application of LCSH and LCGFT; and others due to the limitations of the LC vocabularies themselves. With respect to the MNCAT sub-sample, there was complementarity across the full range of materials: the various forms of literary work were all assigned useful tags not covered by the library subject headings. There were more tags usefully complementing fiction, but mainly because there was more fiction, and more tags assigned to fiction.

Conclusion

It is clear that some library catalogers are adding more headings for fiction and other literary works than was the custom in the past, and that this would make a difference to access even in comparison with the situation DeZelar-Tiedman reported a decade ago. Further, while many literary works found in library catalogs are being entered and tagged in social cataloging platforms such as LT, this tagging is at least as uneven, and probably more so in some ways, than
the indexing being added to bibliographic records. Nevertheless, while library supplied-cataloging appears to be catching up with social cataloging with respect to access provision for belles-lettres, for literary works there are as many subject-related tags available on platforms such as LT as there are headings and subdivisions in academic and public library catalogs. Not all of these tags are necessarily useful: some may be the same as the library terms, some may be inaccurate, and others may represent idiosyncratic views. However, this study confirms that many, perhaps even a majority, could enhance access, and no doubt do within the social cataloging platform.

The question remains how library catalogs can best harness this added value offered by social cataloging, noting that around half of the LT tags do not relate to subject or genre, and some of those that do might also be considered as “noise.” If tags need to be curated, that should probably happen after acquisition and initial processing, when there is more likely to be tags from multiple social catalogers available. How realistic and scalable such curation could be is another matter.

The similar distributions of tag categories for works in the public and university library collections point to the potential utility of social cataloging in different library contexts, even if differences in collection currency, breadth, etc., exist. Tagging behavior and outcomes can still be similar across different materials within a broad area, such as literary works, at least across similar platforms.

This study also highlights particular ways in which social cataloging complements library cataloging with respect to belles-lettres. About a third of non-matching, subject-related concepts pertained to genre and form, with smaller proportions, but significant ones, relating to setting and theme. These aspects are among those included in received models of fiction access, as outlined in the literature review, and likewise feature in the two earlier studies of the social cataloging of fiction by Šauperl, and Pecoskie, Spitteri, and Tarulli. However, the prevalence of these elements in the authors’ analysis suggests that not all of the various elements included in the received models are of equal importance, and that catalogers might do well to focus on a few key elements, such as genre, setting, theme, characters, and authorial attributes, rather than shoehorn their current practices into a long list. In fact, these key elements are generally already aligned with the more modern library cataloging practices. Indeed, Pecoskie, Spitteri, and Tarulli found that genre, location, and period featured proportionately more in headings than tags. Thus it is perhaps less a question of librarians needing to adopt a new framework and more one of integrating the social cataloging into their search systems, leveraging the taggers’ breadth of views and proximity to the material. Further, the existing models omit one element that was covered significantly by the tags examined in this study, namely, authorial attributes, confirming Šauperl’s finding.

Many of the useful non-matching concepts are represented in LCSH and LCGFT; they just were not represented in the headings assigned to the records. This would be partly due to greater numbers of taggers than catalogers and partly to the limits set by general and local cataloging policy. It may also be due to catalogers’ continued emphasis in their daily practice, on subjects, at the expense of genres and settings, both of which overlap with subjects, but can be harder to pin down. This leads to another reason for the additional tags: social catalogers tend to be closer to the work, giving rise to additional knowledge and insights about the work, as well as a stronger impetus, perhaps, to express their views of the work.

Some of the useful non-matching concepts could not be translated into LC terms, however, illustrating the shortcomings of as extensive a controlled vocabulary as LCSH. By its nature, vocabulary control will inevitably lead to a certain loss of expression. Nevertheless, it would be instructive to consider each case that social cataloging raises: in some, the term may in fact be a candidate for inclusion, whereas in others its lack of fit may shed light on the character or structure of the controlled vocabulary.

Some of the tags were particularly valuable because of the minimal cataloging in the corresponding bibliographic record. While library supplied cataloging addresses the need for greater access to belles-lettres and other works of the imagination, it is does not yet do so universally, and more attention required in its coverage of concepts such as genre and setting, which overlap, but go beyond that of “subject,” whether connotatively or denotatively. Even assuming that progress continues, it looks as though there is still potential for social tagging to complement the access provided by library cataloging.

References


2. Christine DeZelar-Tiedman, “Exploring User-Contributed Metadata’s Potential to Enhance Access to Literary Works:


12. Rafferty, “Tagging.”


40. Šauperl, “Four Views of a Novel.”