

Spring 2023

Metadata Decentralized: An Examination of Folksonomy in Book Publishing

Dani Tellvik
Portland State University

Follow this and additional works at: https://pdxscholar.library.pdx.edu/eng_bookpubpaper



Part of the [English Language and Literature Commons](#), and the [Publishing Commons](#)

Let us know how access to this document benefits you.

Recommended Citation

Tellvik, Dani, "Metadata Decentralized: An Examination of Folksonomy in Book Publishing" (2023). *Book Publishing Final Research Paper*. 77.

https://pdxscholar.library.pdx.edu/eng_bookpubpaper/77

This Paper is brought to you for free and open access. It has been accepted for inclusion in Book Publishing Final Research Paper by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

Metadata Decentralized:
An Examination of Folksonomy in Book Publishing
Dani Tellvik
Master of Arts in Book Publishing
Portland State University

Introduction

Classification of materials as multifaceted and complex as books is fraught with competing needs and idiosyncratic forms of expression and organization. A publisher's primary goal is to position their books in a way that will attract readers who find value in the books. Book Industry Standards and Communication (BISAC) codes, the list of industry-approved subject headings, are essential to a book's metadata for North American publishers. They are used to strike the balance so that a book blends in to an established category while also standing out as a particularly compelling or unique expression or the chosen classification.

Metadata Mayhem

Metadata is an invaluable tool for a publisher to get information about their book out into the world ahead of the book's release. However, data doesn't get pushed out or picked up in a uniform way, so a book's metadata can (and often does) become unruly once it's out in the digital space. At Ooligan Press, we ingest a title into our title management system almost a full year before the book is published, and our editorial department applies with Library of Congress for Cataloging-in-Publication (CIP) data well before our sales representatives have given any significant feedback on our metadata. Ooligan's digital department enters metadata into Kindle Direct Publishing separately for our e-books, and this constitutes another entry point for potential inconsistencies.

In 2023, Ooligan faced significant metadata challenges with our winter title. Most online retailers seemed to be picking up data from our title management system, which we checked and updated frequently, but some older versions of our main description were cropping up in library

catalogs. It took weeks to trace these errors back to their source, but we eventually discovered that the old description was coming from the Library of Congress. We had changed the main description seven months prior to the publication date, based on feedback from our sales representatives and our author. Unfortunately, our CIP application had already been submitted before we made the change and the outdated description had distributed to any systems that picked up data from Library of Congress. Our managing editor submitted a correction to LoC, but the book had already been published before the correction was widely distributed.

Another issue we had was with the subtitle. We added the subtitle to the metadata late, and this caused Amazon to re-categorize the book in a completely different genre. As of this writing, six months after the error appeared, we haven't been able to get it resolved. Because Amazon maps publisher-provided metadata to its own internal SEO, we had no way of knowing this would happen and no way to undo it once it was done.

These challenges represent a fraction of the metadata mayhem we—and many publishers of all sizes—face. Because of the unpredictable nature of metadata distribution, there's only so much a publisher can do to ensure accurate categorization and positioning of their books. With the advent of social networking sites centered on reading, however, readers are in a unique position to create and share metadata for their favorite books alongside publishers. While this doesn't address all metadata challenges, it does open up a unique avenue for publishers and authors to get real-time feedback on their positioning efforts.

Social Tagging

Social tagging, the act of applying public, freeform descriptors to digital artifacts, represents reader engagement with the books that have been so carefully positioned by publishers. When

social tags become dense enough to form a collection, they can be referred to as a folksonomy, an imperfect classification system defined by end users. The potential benefit to searchability with the advent of these tags has been studied over the past two decades, but most of the research has focused on tags as they pertain to either library collections or fan fiction repositories. My research examines how readers are redefining book classification post-pub by comparing social tags on Goodreads to the 2022 list of BISAC codes with the aim of observing relevant tag patterns and gaps in market positioning. I hypothesized that tagging behavior would reveal refined genre descriptions and latent interests that may have been overlooked (or unavailable) when BISAC codes were assigned.

A secondary aim of this research is to explore what other types of book information are captured in tags that may be difficult to express with BISAC codes alone. My research suggests that not only are refined genre descriptors apparent in Goodreads tags, but that tags also reveal other dimensions of a book, such as reader behavior and engagement and extra-textual information about the author and the book's connection to current events.

Background

Formal classification is collectively referred to as a taxonomy, a system of categorization to catalog a group of things. In general, taxonomies (also known as *controlled vocabularies*) follow a set of established guidelines, and “[record] the hierarchical and affinitive/ associative relations of a concept,” and “[establish] the size and scope of each topic.” There are a number of taxonomies in use for published materials: BISAC codes, Dewey Decimal Classification (DDC), Library of Congress Subject Headings (LCSH), Machine-Readable-Cataloging records (MARC), Amazon

keywords, and niche organizational systems within bookstores. Though there are different goals for different systems, they all exist to help with content searchability and discovery.

Past research has done a thorough job of exploring tagging and its relationship to social networking sites and libraries. Most sources agree that tagging, in combination with traditional indexing practices, has the potential to create a more complex and inclusive picture of an item. In the literature review, we'll explore two areas of foundational research on social tagging: readers' motivations for using tags, and what information social tags can capture that taxonomies exclude.

Literature Review

Reader Behavior and Engagement

Ownership in the Digital Age

Just as a publisher thinks about how a title fits into their backlist and established brand, so too do readers think about how the books they're interested in fit into their reading interests and goals.

In their 2019 article on literary consumption in the digital age, Anne-Mette Bech Albrechtslund suggests that ownership and cataloging books has changed significantly since the rise of Amazon and Goodreads.¹ They further posit that a reader's sense of ownership extends beyond what they can purchase and that their collection is much more expansive as a result. Goodreads, as the "largest social book cataloging site in the world," provides readers with the opportunity to create a living digital archive of their reading.² When readers apply tags (or shelves, as Goodreads calls them), a book isn't confined to one place on their digital bookshelf; multiple shelves create a complex, multidimensional picture not only of the book's metadata, but of the reader's

1 Anne-Bette Bech Albrechtslund, "Amazon, Kindle, and Goodreads: Implications for Literary Consumption in the Digital Age," *Consumption Markets & Culture* 23, no. 6 (November 1, 2020): 557, <https://doi.org/10.1080/10253866.2019.1640216>.

2 Albrechtslund, "Amazon, Kindle, and Goodreads," 562.

relationship with the book. This kind of categorization is fluid; books can move in and out of collections easily as the reader's relationship to or perception of a book changes.

Motivations for Tagging

In a 2006 study of tagging practices on the bookmarking site del.icio.us, Margaret Kipp and D. Grant Campbell found that users, authors, and indexers had significantly different opinions about where an article belonged.³ Social tagging can't replicate a professional taxonomy because it isn't constrained by rules about how to organize information; however, Kipp and Campbell argue that there's value in this because it can create additional access points to a document.

In 2009, Peishan Bartley built on these ideas by exploring the motivations of LibraryThing users when they tagged books. Bartley chose random participants, had them all tag the same three books, and then recorded explanations about their process. The participants were allowed to create their own tags, but they were asked to choose from predetermined categories of tags: factual, opinion, and personal note. While the tagging environment was manufactured, the study provides a framework through which we can understand how motivations for tagging differ from the goals of traditional cataloging.⁴

Bartley found that 77% of participants cited "collection management" as their main motivation for tagging, and 87% also cited "collection management" as their motivation for using tags provided by other users and already assigned to resources.⁵ Furthermore, Bartley found

3 Margaret Kipp and D. Grant Campbell, "Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices," *Proceedings of the American Society for Information Science and Technology* 43, no. 1 (2006): 14, <https://onlinelibrary.wiley.com/doi/abs/10.1002/meet.14504301178>

4 Peishan Bartley, "Book Tagging on LibraryThing: How, Why, and What Are in the Tags?: Book Tagging on LibraryThing: How, Why, and What Are in the Tags?" *Proceedings of the American Society for Information Science and Technology* 46, no. 1 (2009), <https://doi.org/10.1002/meet.2009.1450460228>.

5 Bartley, "Book Tagging on LibraryThing," 11.

that, although taggers did report being motivated to help others find books, personal collection management was the main reason that participants applied tags by a wide margin.⁶ This is consistent with Albrechtslund's assertion that a reader's sense of ownership extends to the books in their digital collection.

Tags as Collection Enrichment

Kipp and Campbell suggest that collection management tags are valuable for the following reasons:

1. They express a response from the user rather than a statement of the aboutness of the document.
2. They suggest an active engagement with the text, in which the user is linking the perceived subject matter with a specific task or set of interests.⁷

Bartley backs up these claims in their research. When they compared LibraryThing tags to MARC records, they found that even within the 'factual tag' category, which recorded information about the content, physical traits, or author of a book, overlap between tags and MARC was low. Bartley concluded that "tags add to the richness of bibliographic record by enhancing professional records with user reflections and insights that are up to date and speak to current culture."⁸ They saw readers adding complexity to records, rather than obfuscating the categorization that had already taken place.

The research of Marilese Thomas, Dana Caudle, and Cecilia Schmitz further aligns with the assertion that social tagging enhances existing taxonomies. In their study of over 8,000 tags

6 Bartley, "Book Tagging on LibraryThing," 11.

7 Kipp and Campbell, "Patterns and Inconsistencies," 15.

8 Bartley, "Book Tagging on LibraryThing," 16.

for ten books across six library systems (including LibraryThing), they found that 62% of tags added descriptors not captured by assigned LCSH, and only 14% overlapped with LCSH. They also proposed a hybrid system called a *collabulary*, or a compromise between folksonomy and taxonomy systems, wherein a “team of classification experts collaborates with content consumers to create rich, but more systematic content tagging systems.”⁹ This was based on their observation that readers’ familiarity with LCSH seemed to influence some of their tagging practices, which supported their assertion that a hybrid system creates richer metadata for resources.¹⁰

The work done by Kipp and Campbell, Bartley, and Thomas et al. provides a foundational context for tagging practices and how they have interacted with established taxonomies; however, recent research still supports these foundational findings. In their 2021 study, “Tags, Borders, and Catalogs: Social Re-Working of Genre on LibraryThing,” Maria Antoniak, Melanie Walsh, and David Mimno similarly found that social tagging expanded narrow definitions of established genres that had been “mediated by academic scholarship and commercial publishers,” hinting at the value of democratizing book classification in light of the rise of online communities of readers.¹¹

Formal Classification Systems in Book Publishing

In the North America, BISAC codes are the standard system of classification for trade book publishing. The codes are decided upon by Book Industry Study Group (BISG), a committee

9 “Folksonomy,” Wikipedia, accessed November 21, 2008, <http://en.wikipedia.org/wiki/Folksonomies>, quoted in Marilese Thomas, Dana Caudle, and Cecilia Schmitz, “To Tag or Not to Tag?” *Library Hi Tech* 27, no. 3 (November 2009): 411, <https://www.emerald.com/insight/content/doi/10.1108/07378830910988540/full/html>

10 Marilese Thomas, Dana Caudle, and Cecilia Schmitz, “To Tag or Not to Tag?” *Library Hi Tech* 27, no. 3 (November 2009): 425, <https://www.emerald.com/insight/content/doi/10.1108/07378830910988540/full/html>

11 Maria Antoniak, Melanie Walsh, and David Mimno, “Tags, Borders, and Catalogs: Social Re-Working of Genre on LibraryThing,” *Proceedings of the ACM on Human-Computer Interaction* 5, no. CSCW1 (April 13, 2021): 1, <https://doi.org/10.1145/3449103>.

of industry professionals. Once per year, additions, deletions, or changes to the list follow a formal process before they are accepted by the committee.¹² BISAC codes are meant to “facilitate consumer discoverability and ensure usability for book retailers,” according to Tom Richardson of BookNet Canada.¹³ Laura Dawson, a book supply chain consultant for *Publishers Weekly*, writes of BISAC codes, “The BISAC taxonomy is the product of enormous cooperation [between] publishers, booksellers, wholesalers, and other book industry stakeholders.”¹⁴ However, while most taxonomies provide guidelines for how categories should be applied, BISAC remains purposefully undefined. BISG states, “As a general rule, the subjects are not defined. The Committee attempts to create clear and succinct subject descriptors that are not duplicative within the list.”¹⁵ BISAC codes also differ from library classification in that BISG merely provides the list of subject headings, but does not participate in the work of selecting codes for specific works, deferring instead to the publisher’s familiarity with their own titles. The scope of this paper does not permit an exploration of how different publishers may use the same codes in different ways, so I will stop with the implication that BISG leaves room for publishers to apply BISAC codes in ways that may resonate with their specific audiences. That said, publishing is often a self-referential process where industry professionals build their future strategies on what has worked (or not worked) in the past, so there is likely some overarching sense of uniformity in how the codes are applied.

12 “What About Adding New Codes,” Selecting a BISAC Code, BISG, accessed May 5, 2023, <https://www.bisg.org/selecting-a-BISAC-code>

13 Tom Richardson, “The whys of main subjects in Thema and BISAC,” BookNet Canada, February 16, 2021, <https://www.booknetcanada.ca/blog/2021/2/16/the-whys-of-main-subjects-in-thema-and-bisac>

14 Laura Dawson, “On the Tyranny of Keywords,” *Publishers Weekly*, April 14, 2017. <https://www.publishersweekly.com/pw/by-topic/industry-news/bookselling/article/73341-on-the-tyranny-of-keywords.html>

15 “Are There Definitions for the Various Subject Headings,” General BISAC Questions, BISG, accessed May 5, 2023, <https://www.bisg.org/general-BISAC-questions>

It's difficult to overstate how pervasive BISAC codes are in trade publishing. Most North American businesses require BISAC codes to be assigned for a book to be distributed and carried.¹⁶ BISG recommends a limit of three BISAC codes per book, and encourages publishers to identify a “main” subject in their metadata.¹⁷ Richardson explains that the main subject is a “focal point for your sales effort,” because it affords retail partners the convenience of knowing what a book is mostly about at a glance.¹⁸ Library classification of books is also based on BISAC codes to some degree. For the CIP program, Library of Congress catalogers base their classification on publisher-provided metadata.¹⁹ Through BISAC codes, publishers assert a great deal of influence over a book's classification trajectory and therefore which audiences a book is visible to.

The greatest strength of the BISAC subject headings—and most controlled vocabularies—are their ability to describe the relationship between concepts. The codes loosely organize language in a way that maps “multiple explicit relationships between terms, which can be broader, narrower, or related to each other.”²⁰ So, there's a difference between Fiction: Nature & Environment; Nature: Environmentalism; and Environmentalism. These differences are inherent in the categorization system whereas a reader-applied tag might just say ‘Nature’ without indicating whether a resource is fiction or nonfiction.

16 “Who Uses the BISAC Subject Headings?” BISAC FAQ, BISG, accessed May 1, 2023, <https://www.bisg.org/BISAC-FAQ>

17 “What if my Book has More than One Subject Matter,” BISAC FAQ, BISG, accessed April 27, 2023, <https://www.bisg.org/BISAC-FAQ>

18 Richardson, “Main Subjects,” February 6, 2021.

19 “Purpose of the CIP Program,” About CIP, Library of Congress, accessed May 11, 2023, <https://www.loc.gov/publish/cip/about/>

20 Vocabulary Control,” Updated March 16, 2020.

Folksonomies in Book Publishing

As covered in the literature review, readers use their own classification systems to actively participate in a book's metadata footprint when they assign tags not just in spite of the fact that their systems are idiosyncratic but because of it. Unlike publishers and librarians, readers have no obligation to think about how others will access their collections. Sites like Goodreads and LibraryThing not only make these private systems of classification public and visible; they also allow them to contribute to the larger language of reader-generated metadata.

Reader-generated metadata falls under the umbrella of folksonomy. A folksonomy differs from a taxonomy in that it is not overseen or controlled by shared guidelines or committee. As of this writing, one of the most comprehensive definitions of folksonomy is the one provided by Wikipedia. Thomas et al. used the 2009 Wikipedia entry in their foundational study:

Folksonomy (also known as collaborative tagging, social classification, social indexing, and social tagging) is the practice of collaboratively creating and managing tags to annotate and categorize content. Folksonomy describes the bottom-up classification system that emerges from social tagging. In contrast to traditional subject indexing, metadata is generated not only by experts but by creators and consumers of the content. Usually, freely chosen keywords are used instead of a controlled vocabulary.²¹

Since then, the definition has changed in a few key ways. As tagging practices have become more complex, we have begun to recognize different types of folksonomies. We have also wrapped user motivations into conversations about folksonomy. The 2023 Wikipedia entry reflects this:

Folksonomy is a classification system in which end users apply public tags to online items, typically to make those items easier for themselves or others to find later. Over time, this can give rise to a classification system based on those tags and how often they are applied

21 "Folksonomy," Wikipedia, accessed November 21, 2008, <http://en.wikipedia.org/wiki/Folksonomies>, quoted in Marilese Thomas, Dana Caudle, and Cecilia Schmitz, "To Tag or Not to Tag?" *Library Hi Tech* 27, no. 3 (November 2009): 411, <https://www.emerald.com/insight/content/doi/10.1108/07378830910988540/full/html>

or searched for, in contrast to taxonomic classification designed by the owner of the content and specified when it is published.²²

In the 2023 entry, ‘social tagging’ and ‘collaborative tagging’ are untangled from the main definition as separate terms. The difference between the two is in how much the users are collaborating to create a shared vocabulary. Social tagging relies less on intentional tagging and is more focused on individual methodologies. Goodreads and LibraryThing are examples of social tagging. The popular fan fiction site, Archive of Our Own (AO3), is an example of collaborative tagging where content creators and their readers are in conversation with each other, and tags often take on stable meanings over time. AO3 has an additional layer of collaboration because volunteers called metadata wranglers work to keep the folksonomy clean and easy to search.²³

Strengths of Folksonomies

Scholars agree that there are four main strengths of folksonomies:

1. The ability to capture the complex nature of a subject with the input of so many different perspectives
2. The collaborative nature of tagging and low barrier to entry
3. The flexibility with which users can assign labels using their own language and ideas
4. The speed with which folksonomies can respond to changes in language usage

I would elevate a fifth dimension that only a few sources touch upon: tags allow more serendipitous discovery than sorting by genre alone. On sites like Goodreads, if you click on a tag that describes a particularly niche element of the book you’re looking at, it can automatically connect you to other books with that same niche element. In this way, folksonomies can be

²² “Folksonomy,” Wikipedia, last modified May 7, 2023, <https://en.wikipedia.org/wiki/Folksonomy>

²³ “Wrangling Guidelines: Intro and General Concepts,” Archive of Our Own, accessed November 8, 2022. https://archiveofourown.org/wrangling_guidelines/2.

used for more precise filtering than the tools usually available to readers on retailer websites and library catalogs.

The other folksonomy strengths that are especially applicable to this research are freedom of expression and flexibility. In their study of folksonomies and information retrieval, Isabella Peters and Wolfgang Stock suggest that folksonomies decentralize classification by allowing users to “authentically represent the language of authors and users.”²⁴ Readers, therefore, aren’t restricted to a taxonomy that may be unfamiliar or even exclusive.

Peters and Stock further observe that modifications, updates, new terminology, and deletion of old terminology happen much more quickly within a folksonomy because of the low barrier to entry for users. They write, “tags guarantee a fast response to changes and innovations in the knowledge domain.”²⁵ Tags can be applied, changed, or deleted as many times as the reader deems necessary. This degree of flexibility allows readers to quickly change tags that may use outdated language without having to go through a committee like BISG, which takes months to deliberate.

Limitations of Folksonomies

While folksonomies allow for a high degree of collaboration and flexibility, they can become disorganized, chaotic, and difficult to navigate. In a folksonomy, there is no way to “indicate broader, narrower, or related terms and their relationships and no rules for determining the desired level of specificity.”²⁶ Additionally, synonymous words are treated as different tags and won’t appear in searches for similar words unless tag-wrangling has been done. The same is true

24 Isabella Peters and Wolfgang G. Stock, “Folksonomy and Information Retrieval,” *Proceedings of the American Society for Information Science and Technology* 44, no. 1 (2007): 13, <https://doi.org/10.1002/meet.1450440226>

25 Peters and Stock, “Folksonomy and Information Retrieval,” 13.

26 Thomas, Caudle, and Schmitz, “To Tag or Not to Tag?” 414.

of variations: singular/plural, past/present tense, hyphenated/unhyphenated, etc. Peters and Stock identify the conflation of different dimensions of a description as another limitation of folksonomies. They caution that tagging can lead to “the problem that different levels of ofness, aboutness, iconology, isness melt into one single level.”²⁷ This makes folksonomy behaviors difficult to observe and measure.

Comparison of Controlled and Uncontrolled Vocabularies

	BISAC Subject Headings	Folksonomies / Social tagging
Strengths	<ul style="list-style-type: none"> • Curated by industry professionals who have deep knowledge of market trends and discuss alterations and additions as a committee • Preserves and highlights relationships between terms • Helps with positioning when marketing books 	<ul style="list-style-type: none"> • No limit on how many tags can be added to a specific work • Responds quickly to changes in language usage • Readers can participate in the classification of the content they're reading • Represents complexities • Readers can describe books in their own language
Limitations	<ul style="list-style-type: none"> • Slow to update/change • Not updated post-pub even if more specific codes are added • Cannot capture all aspects of a book • The way codes are being shaped may leave perspectives out of the conversation, leaving to more bias and lack of representations 	<ul style="list-style-type: none"> • Doesn't make visible relationships or hierarchy between tags • Some tags will always remain obscure or ambiguous • No uniformity in tag application

Figure 1: A comparison of the strengths and limitations of BISAC codes and folksonomies discussed in this paper.

²⁷ Peters and Stock, “Folksonomy and Information Retrieval,” 16.

Application of Tagging in Trade Publishing

While extensive research has been done to support the validity of tagging in library classification, there has been less attention paid to how tags interact with BISAC codes. What my research has ultimately shown is that there isn't a useful way to implement a tagging system within trade publishing in its current state and that it would, in fact, be contrary to the intent of social tagging as reader-generated classification. In this section, I discuss where tag-like strategies have been used in trade publishing before and speculate on why it hasn't been successful.

The Problem with Keywords

Currently, keywords are the closest system to tagging that publishers have at their disposal. Keywords can be added to metadata as a supplement to BISAC codes for increased discoverability. They do have some limitations, however. For one, they're often dependent on how well a book is selling. Laura Dawson disowns keywords as "unruly," and claims that "if your book doesn't sell more copies within a day or so after keyword optimization, it's relegated to the last pages of search results."²⁸ This is echoed by Chris Sims, the founder of Kadaxis—a company that helps publishers boost the effectiveness of their metadata—who states that keywords don't really help with discoverability unless your book appears on the first page or results and/or maintains steady sales.²⁹ This means keyword effectiveness for small publishers is limited. To compound this, some retailers don't use publisher-supplied keywords, including the large bookstore chain, Barnes & Noble.³⁰ Furthermore, the one retailer guaranteed to take publisher-provided metadata,

28 Dawson, "Tyranny of Keywords," April 14, 2017.

29 Chris Sims, "Measuring Keyword Effectiveness on Amazon," Kadaxis, January 8, 2018, <http://kadaxis.com/blog/2018/1/8/measuring-keyword-effectiveness-on-amazon> Kadax

30 "Keywords for Searching Provided by Publishers and Authors," Barnes & Noble, accessed October 20, 2022, <https://help.barnesandnoble.com/hc/en-us/articles/5356889970715-Keywords-for-Searching-Provided-by-Publishers-and-Authors>

Amazon, maps keywords to their own internal system. As demonstrated by the anecdotes at the beginning of this paper, their mapping is imperfect and can lead to books being miscategorized.

Angry Robot: An Informal Case Study

In a blog post entitled, “Everything is Miscellaneous: Why Publishing Needs Tagging,” Michael Underwood, then employed at London-based publisher Angry Robot, suggests publishers could use tags to combat the restrictions of genres, pointing to the potential for tags to broaden the definition and classification of a particular book rather than narrowing it. He further asserts that tags could “round out the work done by the title, cover art, design, and cover copy.”³¹

Underwood’s speculations, though not captured in formal research, align with the studies previously mentioned. His assertion that tags build upon the positioning work that publishers have already done reveals that publishing professionals (myself included as the outset of this research) tend to think of tags as further outward-facing positioning when that’s not the motivation for most readers. This points to a discrepancy between the potential publishers see in tagging and the reality of how tags are currently used. If we compare Underwood’s proposed tags for *Pride and Prejudice* to the top Goodreads tags, this becomes even clearer.



And looking back at a classic, how about *Pride and Prejudice*, by Jane Austen:
#Romance #Adult #Regency #Hate At First Sight #Suitors #Order of
Marriage #Sisters #England #Gentry #Banter #Drawing Room Politics
#Gender

Figure 2: Underwood’s proposed tags of *Pride and Prejudice*

31 Michael Underwood, “Everything is Miscellaneous: Why Book Publishing Needs Tagging,” Boingboing, November 20, 2015, <https://boingboing.net/2015/11/20/everything-is-miscellaneous-w.html>

Top shelves for <i>Pride and Prejudice</i> Showing 1-100 of 95,207					
to-read	2,096,225 people	school	1,349 people	adult-fiction	774 people
currently-reading	197,566 people	my-books	1,321 people	wishlist	769 people
classics	64,902 people	2018	1,318 people	2014	757 people
favorites	49,242 people	19th-century	1,258 people	england	747 people
fiction	18,231 people	2016	1,254 people	1001-books	701 people
romance	17,064 people	book-club	1,225 people	audible	692 people
classic	10,011 people	5-stars	1,210 people	regency	690 people
		2017	1,193 people	2013	683 people

Figure 3: Goodreads tags added to *Pride and Prejudice*, May 2023

The tags that describe the “aboutness” of the book do have some overlap (England, Romance, Adult, Regency); however, there are other dimensions that are entirely unconsidered in Underwood’s tags, while Goodreads users didn’t identify some of the more niche elements that Underwood highlights.

Angry Robot experimented with placing suggested tags within their main descriptions to help readers identify subgenres while browsing. There is no indication of how successful this practice was, only the absence of such tags in the publisher’s more recent releases. This could suggest that readers didn’t necessarily take to the cheeky tagging system thrust upon them by the publisher. Tags provided by the publisher run into the same limitations that other publisher-provided metadata does. It’s not collaborative or flexible, and it is often only used to further the publisher’s marketing efforts without leaving room for feedback from the readers themselves.

Methodology

I designed my research in the spirit of Thomas et al.’s comparison of tags with the Library of Congress Subject Headings, but chose to look at BISAC codes instead. I combined a case study approach with content analysis, taking into consideration the size of the publisher and popularity

of the book, both of which matter for tag density, or the number of tags applied to a book. I narrowed my focus to small- to mid-size independent publishers (not imprints) because I believe the results of this study will be most actionable to them. Ultimately, I chose the Portland-based publisher Tin House, which publishes between fifteen and twenty titles per year, making them a sizable small publisher with ample tag data on Goodreads. I chose books published in 2021 based on the assumption that readers would have been more active online due to COVID-19 restrictions that made shopping at brick and mortar bookstores more difficult.

Goodreads is an example of what I would call a *wild folksonomy*, in that the tags are completely reader-driven; no outside catalogers are coming in to clear up the tags or reconcile duplicates. Since the goal of my research wasn't to observe *what* tags were being applied to books but to pick up any significant patterns in these tags, I needed a greater number of tags rather than a greater number of books. I would have been able to study more books on LibraryThing because there are fewer tags per book, but on Goodreads, I had a denser selection of tags.

Each book I chose had between 300 and 2,100 tags (with the exception of three poetry books that had only a handful of tags each), which would have been too time-consuming to sift through for the purposes of this research. The compromise I came up with was taking a sample of the first 200 tags per book. Goodreads sorts their shelves based on frequency, so the further down you get on the list of shelves, the more you find tags that are either misspellings of tags that have already been used or tags that are so individualized as to be virtually meaningless except to the person who created them.

Once I had collected this data, I did tag-wrangling to reconcile alternate spellings, plural/singular differences, and, in a few cases, synonyms. For example, nonfiction can be represented in

a number of different ways that each appear as a different tag: non-fiction, nonfiction, nonfic, nf, etc. I tried to have as light a touch as possible with the tags so as not to guess at and thereby warp any intended meanings. If there was a case I wasn't sure about, I left it as a separate tag.

I organized the tags into four different categories: describing reader relationship to a book, categorizing the content of the book itself, a combination of relationship and subject classification (for example, fiction-to-be-read), and miscellaneous.

Discussion

Tags are extremely difficult to capture in clearly delineated categories. Because each reader applied tags based on their own unique language, it was often impossible to grasp the intention behind any given tag. For example, does '2021-reads' mean the same as 'read-in-2021'? Or does 2021-reads indicate that the reader is planning to read the book in 2021? Or does it mean that the book was published in 2021? Social tagging systems are co-created, but consumers are often working in parallel to each other, rather than in conjunction. This means that even though two users might select the same tag, there's no way to know if the tag means the same thing to each user. Each tag likely has multiple meanings.

To return to Peters and Stock's observation that tags are often conflated across dimension of meaning, I further add that this may lead to dilution in collections bound by the same tag. Goodreads may pull together books that have only tangential relationships to a tag. For example, for the book *Always Crashing in the Same Car*, the tag 'cancer' is applied, but it's unclear to what degree the book is *about* cancer. This observation further contributes to the assertion that a collaborative effort between taxonomists and folksonomists would produce metadata that is more comprehensive while still maintaining coherence.

Even though getting quantitative, granular data about subject classification wasn't as successful as I'd hoped, there are broader observations that could be useful for trade book publishing practices. Furthermore, the difficulty in quantifying tidy categories for tags on Goodreads speaks to the definition of a folksonomy; Goodreads tags are totally driven by readers describing books in their own words. As such, I pulled the following patterns and findings out of the data I collected:

1. My research aligns with previous research conducted within libraries: tags do add layers of specificity to genre and subject matter descriptions
2. The overwhelming majority of tags applied on Goodreads referenced personal collection management
3. Tags about authors suggest that readers are interested in finding and elevating underrepresented authors

Below are the graphs that capture the averages for each genre I studied: Fiction, Nonfiction, and Poetry:

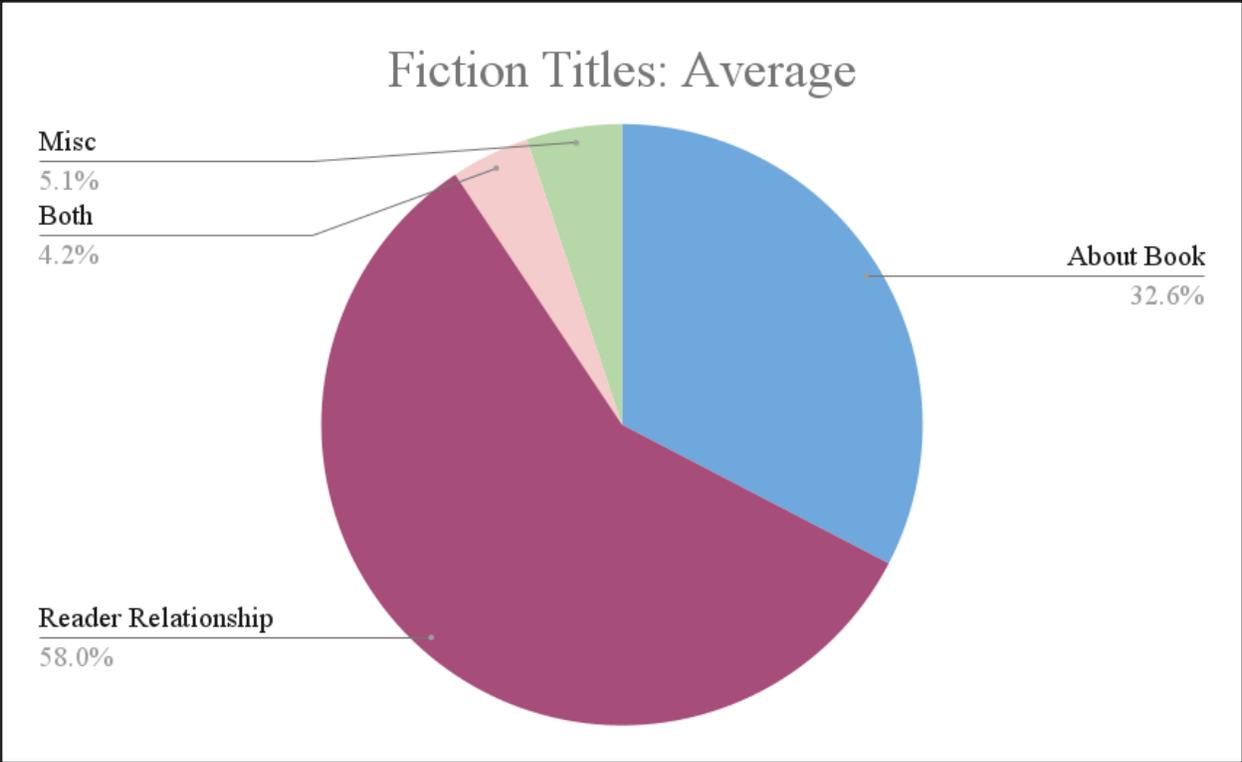


Figure 4: For the fiction titles I included, over half of the tags described a reader’s relationship to the book in terms of task management (to-read, want-to-read), motivation (book clubs, giveaways), and the source or format of the book (audiobook, ebook, library, Libby), and opinion. The next largest category was tags describing details about the book, either further classifying the book by subject or including extra-textual metadata, such as details about the author or publication date.

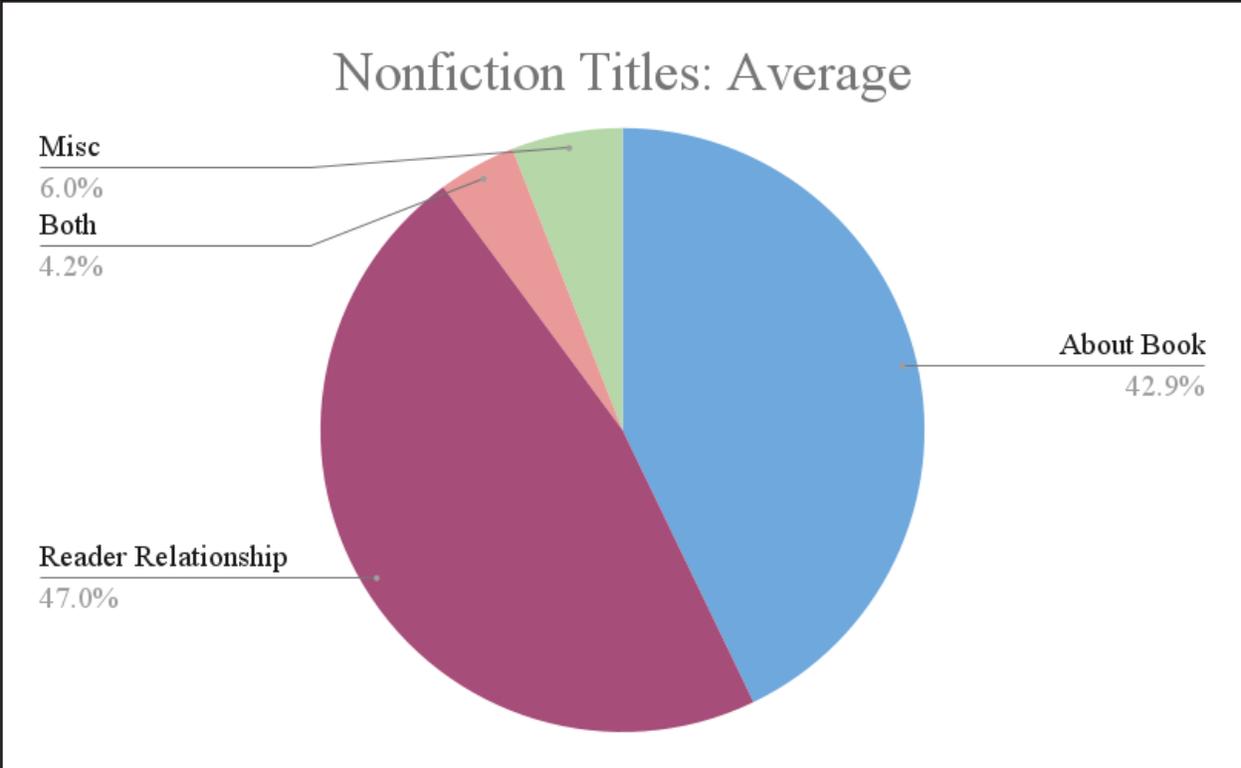


Figure 5: For Tin House’s 2021 nonfiction titles (most of which were biography or autobiography), there was a higher percentage of classification-based tags than reader-relationship tags.

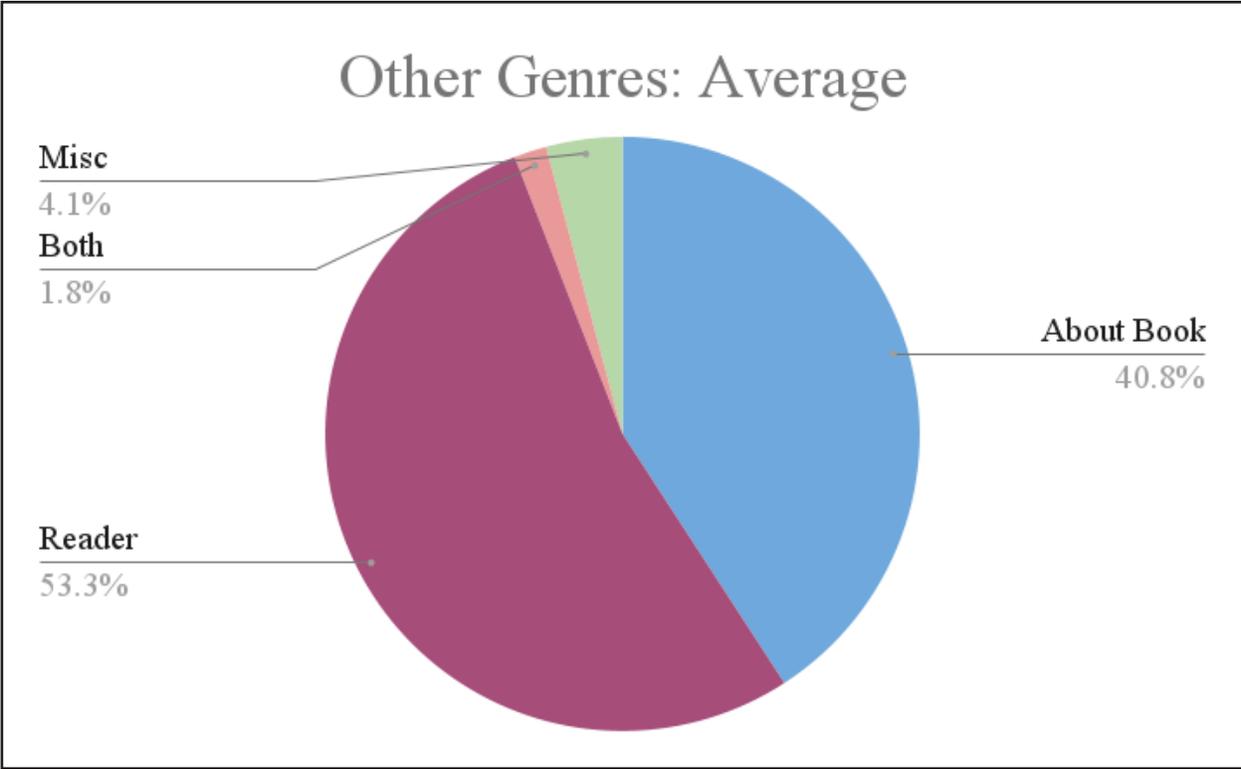


Figure 6: This graph shows the breakdown of Tin House’s 2021 literary collections and poetry books. The profile is most similar to Tin House’s fiction titles.

Limitations

In doing this research, I did not study frequency metrics for tags on Goodreads nor did I delve into the rich subject of algorithms that formulate book recommendations for readers based on data that sites collect from readers' tagging practices. I don't have a background in data analysis or collection. My method for categorizing and organizing tags was imprecise and not reliable enough for more than the broad strokes analysis I conducted in my discussion.

Key Findings

Latent subgenres

Goodreads tags do reveal subgenres that aren't as easily displayed in publisher-provided metadata. One particular example of notes is the title, *Justine*, to which the BISAC code FICTION/Friendship was applied. However, 11% of the tags within the subject classification category explicitly refer to LGBTQ+ themes in the book or categorized it as LGBTQ+. Visibility for LGBTQ+ content is increasing, but tags are already doing the work of elevating latent or implied LGBTQ+ elements, and at a quicker rate.³²

Reader Relationship Tags

While the majority of tags were applied in service of an individual reader's collection management, there was still a lot of rich data that could be pulled from the percentage of tags describing reader-and-book relationships, including task-management (read, owned, want-to-read, etc.), opinion, or inclusion in book clubs or library collections. Another way to interpret this is that readers are asserting "ownership" of books by describing their relationship with them

³² This is also suggested by Hanna Ziegler in her paper "Coding LGBTQ Content: BISACs, Fanfiction, and Searchability in the Digital Age." Full citation can be found in the bibliography.

in tags. Publishers tend to focus on sales, but taking note of how readers are “claiming” books in their online collections may yield other useful metrics by which to gauge reader engagement with a book. For instance, for a book that has multiple book club tags applied, tags may indicate regional popularity that was difficult to measure by sales data alone. Furthermore, social networking sites like Goodreads make it easier for readers to discover and engage with titles after a book is published and the marketing campaign is largely over. Most of a publisher’s effort is focused on the pre-launch phase of a book’s life, but the timelessness of discovery on Goodreads (and other sites like it) makes a strong case for more care and feeding of one’s backlist titles. If publishers look at what tags readers are applying to books in the weeks and months following its release, they might find clues about how to adjust a book’s positioning or work on social media campaigns that highlight aspects of the book they didn’t initially anticipate being popular with readers.

Amplifying Diverse Voices

Codifying an author’s identity in something as formal as BISAC brings up potential risks for the author and the reception of their work. BISAC co-opted the OwnVoices hashtag as a BISAC code for adult and young adult fiction just six months before We Need Diverse Books formally announced that they were no longer going to use the descriptor. We Need Diverse Books stated that the hashtag had become a “catch-all’ marketing term” for the publishing industry and that the vagueness of the term didn’t accurately reflect complex identities or respect an author’s own words about themselves and their characters.³³ The organization also brought up that the term

³³ Alaina Lavoie, “Why We Need Diverse Books is No Longer Using the Term #OwnVoices,” *We Need Diverse Books*, June 6, 2021, <https://diversebooks.org/why-we-need-diverse-books-is-no-longer-using-the-term-ownvoices/>

could “place diverse creators in uncomfortable and potentially unsafe situations.”³⁴

Many readers, however, tag book based on different aspects of the author’s identity with tags like ‘author-of-color,’ ‘black-writers,’ and ‘female-author,’ to name a few. The prevalence of these tags could indicate that readers are eager to do seek out and make visible authors from underrepresented backgrounds, something mainstream publishing has promised to do but has been slow to deliver on. I predict that as the publishing industry works toward amplifying more diverse voices, readers will continue to carry out this work in their tagging practices, allowing the two systems to work in parallel and hopefully gain insight from each other.

Furthermore, tags on Goodreads create their own collections, so if you click on ‘author-of-color,’ you will automatically be brought to another page with other books that readers have tagged ‘author-of-color.’ The increased visibility from these tags could, in turn, be helpful to publishers as they look for comp titles in the acquisition and marketing process to begin addressing the exclusionary comping practices addressed in Laura B. McGrath’s 2019 article, “Comping White.”³⁵ Any further analysis on this point is beyond the scope of this research, but I include it as an avenue publishers could explore as one small part of their efforts to elevate underrepresented voices and change exclusionary practices in publishing. Tagging may carry similar risks to the OwnVoices BISAC code, in terms of not respecting an author’s language about their own identity, however, because of the responsiveness of tagging practices, these risks have a greater chance of being mitigated more quickly and directly.

34 Alaina Lavoie, “Why We Need Diverse Books is No Longer Using the Term #OwnVoices,” June 6, 2021.

35 Laura B. McGrath, “Comping White,” Los Angeles Review of Books, January 21, 2019, <https://lareviewofbooks.org/article/comping-white/>

Suggestions for Further Research

There are many different directions that research could go from here, as the scope of this paper was relatively narrow. More precise tag analysis or tag analysis with a larger sample size is one. Further research could also be conducted in comparing tags to marketing copy to pursue other avenues of using reader-defined language about books to more finely tune metadata practices and marketing efforts.

Bibliography

- Albrechtslund, Anne-Mette Bech. "Amazon, Kindle, and Goodreads: Implications for Literary Consumption in the Digital Age." *Consumption Markets & Culture* 23, no. 6 (November 1, 2020): 553-68. <https://doi.org/10.1080/10253866.2019.1640216>.
- Antoniak, Maria, Melanie Walsh, and David Mimno. "Tags, Borders, and Catalogs: Social Re-Working of Genre on LibraryThing." *Proceedings of the ACM on Human-Computer Interaction* 5, no. CSCW1 (April 13, 2021): 1–29. <https://doi.org/10.1145/3449103>.
- Archive of Our Own. "Wrangling Guidelines: Intro and General Concepts." Accessed November 8, 2022. https://archiveofourown.org/wrangling_guidelines/2.
- Barnes & Noble. "Keywords for Searching Provided by Publishers." Accessed October 20, 2022. <https://help.barnesandnoble.com/hc/en-us/articles/5356889970715-Keywords-for-Searching-Provided-by-Publishers-and-Authors>.
- Bartley, Peishan "Book Tagging on LibraryThing: How, Why, and What Are in the Tags?" *Proceedings of the American Society for Information Science and Technology* 46, no. 1 (2009): 1–22, <https://doi.org/10.1002/meet.2009.1450460228>.
- Book Industry Study Group. "What About Adding New Codes," Selecting a BISAC Code. Accessed May 5, 2023. <https://www.bisg.org/selecting-a-BISAC-code>.
- Book Industry Study Group. "Complete BISAC Subject Headings List, 2021 Edition." Accessed May 12, 2023. <https://bisg.org/page/BISACEdition>.
- Book Industry Study Group. "What if my Book has More than One Subject Matter," BISAC FAQ. Accessed April 27, 2023, <https://www.bisg.org/BISAC-FAQ>.

- Book Industry Study Group. "Who Uses the BISAC Subject Headings?" BISAC FAQ. Accessed May 1, 2023, <https://www.bisg.org/BISAC-FAQ>.
- Dawson, Laura. "On the Tyranny of Keywords." *Publishers Weekly*, April 14, 2017. <https://www.publishersweekly.com/pw/by-topic/industry-news/bookselling/article/73341-on-the-tyranny-of-keywords.html>.
- Gerolimos, Michalis. "Tagging for Libraries: A Review of the Effectiveness of Tagging Systems for Library Catalogs." *Journal of Library Metadata* 13, no. 1 (January 2013): 36-58. <https://doi.org/10.1080/19386389.2013.778730>.
- Kipp, Margaret and D. Grant Campbell. "Patterns and Inconsistencies in Collaborative Tagging Systems: An Examination of Tagging Practices." *Proceedings of the American Society for Information Science and Technology* 43, no. 1 (2006): 1-18. <https://onlinelibrary.wiley.com/doi/abs/10.1002/meet.14504301178>.
- Lavoie, Alaina. "Why We Need Diverse Books is No Longer Using the Term #OwnVoices." *We Need Diverse Books*, June 6, 2021. <https://diversebooks.org/why-we-need-diverse-books-is-no-longer-using-the-term-ownvoices/>
- Librarianship Studies & Information Technology. "Vocabulary Control." Updated March 16, 2020. <https://www.librarianshipstudies.com/2017/03/vocabulary-control.html>.
- Library of Congress. "Purpose of the CIP Program." About CIP. Accessed May 11, 2023, <https://www.loc.gov/publish/cip/about/>.
- McGrath, Laura B. "Comping White." *Los Angeles Review of Books*, January 21, 2019. <https://lareviewofbooks.org/article/comping-white/>.

Peters, Isabella, and Wolfgang G. Stock. "Folksonomy and Information Retrieval." *Proceedings of the American Society for Information Science and Technology* 44, no. 1 (2007): 1–28.

<https://doi.org/10.1002/meet.1450440226>.

Richardson, Tom. "The whys of main subjects in Thema and BISAC." *BookNet Canada*, February 16, 2021. <https://www.booknetcanada.ca/blog/2021/2/16/the-whys-of-main-subjects-in-thema-and-bisac>.

Sims, Chris. "Measuring Keyword Effectiveness on Amazon." Kadaxis, January 8, 2018. <http://kadaxis.com/blog/2018/1/8/measuring-keyword-effectiveness-on-amazon>.

Thomas, Marliese, Dana Caudle, and Cecilia Schmitz. "To Tag or Not to Tag?" *Library Hi Tech* 27, no. 3 (September 4, 2009): 411–34. <https://doi.org/10.1108/07378830910988540>.

Underwood, Michael. "Everything is Miscellaneous: Why Book Publishing Needs Tagging." *Boingboing*, November 20, 2015. <https://boingboing.net/2015/11/20/everything-is-miscellaneous-w.html>.

Wikipedia, "Folksonomy." Last modified May 7, 2023, <https://en.wikipedia.org/wiki/Folksonomy>.

Ziegler, Hanna. "Coding LGBTQ Content: BISACs, Fanfiction, and Searchability in the Digital Age." Book Publishing Final Research Paper, (2019). https://pdxscholar.library.pdx.edu/eng_bookpubpaper/44/?utm_source=pdxscholar.library.pdx.edu%2Feng_bookpubpaper%2F44&utm_medium=PDF&utm_campaign=PDFCoverPages.

Suggestions for Further Reading

Archive of Our Own

A huge foundational area of this research that didn't make it into this paper is the collaborative tagging system featured on Archive of Our Own. The tag-wrangling, like most things on the site, is entirely driven by volunteer labor. Below are a list of articles that explore this subject in more depth than I was able to include in this paper.

Johnson, Shannon Fay. "Fan fiction metadata creation and utilization within fan fiction archives:

Three primary models." *Transformative Works and Cultures* 17, no. 1 (2014).

Lothian, Alexis. "Archival Anarchies: Online Fandom, Subcultural Conservation, and the

Transformative Work of Digital Ephemera," *International Journal of Cultural Studies* 16, no. 6 (November 2013): 541-56. <https://doi.org/10.1177/1367877912459132>.

Price, Ludi. "Fandom, Folksonomies and Creativity: The Case of the Archive of Our Own."

In *The Human Position in an Artificial World: Creativity, Ethics and AI in Knowledge Organization*, edited by David Haynes and Judi Vernau, 11–37. Würzburg: Ergon Verlag, 2019, <https://doi.org/10.5771/9783956505508-11>.

Price, Ludi, and Lyn Robinson. "Tag Analysis as a Tool for Investigating Information Behaviour:

Comparing Fan-Tagging on Tumblr, Archive of Our Own and Etsy." *Journal of Documentation* 77, no. 2 (January 1, 2021): 320–58. <https://doi.org/10.1108/JD-05-2020-0089>.

Research on Algorithms

Dourish, Paul. "Algorithms and Their Others: Algorithmic Culture in Context." *Big Data & Society* 3, no. 2 (2016): 2053951716665128.

Hallinan, Blake, and Ted Striphas. "Recommended for you: The Netflix Prize and the Production of Algorithmic Culture." *New media & society* 18, no. 1 (2016): 117-137.

Murray, Simone. "Secret agents: Algorithmic culture, Goodreads and Datafication of the Contemporary Book World." *European Journal of Cultural Studies* 24, no. 4 (2021): 970-989.

Striphas, Ted. "Algorithmic culture." *European Journal of Cultural Studies* 18, no. 4-5 (2015): 395-412.

Appendix A: Further Bibliographic Information about Books Studied

Title	Publisher-supplied BISAC codes (from Books In Print)	Dataset
What Storm, What Thunder	FIC051000 - FICTION / Cultural Heritage FIC070000 - FICTION / Disaster FIC019000 - FICTION / Literary	Fiction
Always Crashing in the Same Car: On Art, Crisis, and Los Angeles, California	BIO007000 - BIOGRAPHY & AUTOBIOGRAPHY / Literary Figures BIO026000 - BIOGRAPHY & AUTOBIOGRAPHY / Personal Memoirs BIO005000 - BIOGRAPHY & AUTOBIOGRAPHY / Entertainment & Performing Arts	Nonfiction
Superdoom	POE023010 - POETRY / Subjects & Themes / Death, Grief, Loss POE024000 - POETRY / Women Authors POE023020 - POETRY / Subjects & Themes / Love & Erotica	Insufficient number of tags to be included
The Way She Feels	BIO038000 - BIOGRAPHY & AUTOBIOGRAPHY / Survival BIO026000 - BIOGRAPHY & AUTOBIOGRAPHY / Personal Memoirs	Nonfiction
Justine	FIC019000 - FICTION / Literary FIC071000 - FICTION / Friendship FIC043000 - FICTION / Coming of Age	Fiction
Bride of the Sea	FIC019000 - FICTION / Literary FIC045010 - FICTION / Family Life / Marriage & Divorce FIC051000 - FICTION / Cultural Heritage FIC043000 - FICTION / Coming of Age	Fiction
The Kissing Bug: A True Story of a Family, an Insect, and a Nation's Neglect of a Deadly Disease	BIO017000 - BIOGRAPHY & AUTOBIOGRAPHY / Medical (incl. Patients) HEA039110 - HEALTH & FITNESS / Diseases & Conditions / Nervous System (incl. Brain) SOC026030 - SOCIAL SCIENCE / Sociology / Urban	Nonfiction

Something Wonderful	FIC009010 - FICTION / Fantasy / Contemporary FIC019000 - FICTION / Literary FIC029000 - FICTION / Short Stories (single author)	Insufficient number of tags to be included
All the Names Given: Poems	POE023020 - POETRY / Subjects & Themes / Love & Erotica POE023040 - POETRY / Subjects & Themes / Places POE023050 - POETRY / Subjects & Themes / Family	Insufficient number of tags to be included
How to Order the Universe	FIC019000 - FICTION / Literary FIC045000 - FICTION / Family Life / General FIC043000- FICTION / Coming of Age	Fiction
Perseverance	POE005020 - POETRY / European / English, Irish, Scottish, Welsh POE023010 - POETRY / Subjects & Themes / Death, Grief, Loss POE012000 - POETRY / Caribbean & Latin American	Other
White Magic	LCO013000 - LITERARY COLLECTIONS / Indigenous Peoples in the Americas LCO019000 - LITERARY COLLECTIONS / Women Authors LCO010000 - LITERARY COLLECTIONS / Essays	Other
Win Me Something	FIC045000 - FICTION / Family Life / General FIC051000 - FICTION / Cultural Heritage FIC054000 - FICTION / Asian American FIC043000 - FICTION / Coming of Age	Fiction
Unsettled Ground	FIC019000 -FICTION / Literary FIC025000 - FICTION / Psychological FIC045020 - FICTION / Family Life / Siblings	Fiction
Windswept: Walking the Paths of Trailblazing Women	BIO007000 - BIOGRAPHY & AUTOBIOGRAPHY / Literary Figures BIO022000 - BIOGRAPHY & AUTOBIOGRAPHY / Women BIO026000 - BIOGRAPHY & AUTOBIOGRAPHY / Personal Memoirs	Nonfiction

More on Tin House

Tin House is a Portland-based publisher of exemplary books. Here's the description from their website:

Publisher of award of award-winning books of literary fiction, nonfiction, and poetry; home to a renowned workshop and seminar series; and partner of a critically acclaimed podcast, Tin House champions writing that is artful, dynamic, and original. Tin House authors have garnered acclaim everywhere from The New York Times and NPR to The Wall Street Journal and People magazine. We are proud to publish and promote writers who speak to a wide range of experience, lending context and nuance to their examination of our world.¹

¹ "About Us," Tin House, accessed May 5, 2023, <https://tinhouse.com/about-tin-house/>

Appendix B: A Closer Look at the Data

Fiction

	What Storm, What Thunder	Justine	How to Order the Universe	Bride of the Sea	Win Me Something	Unsettled Ground	Average
Task/Behavior	52	37	43	48	64	47	49
Genre/Subject	27	63	32	49	24	29	37
Metadata	19	19	20	15	23	4	17
Source/format	17	8	9	11	10	11	11
Motivation	22	5	9	20	16	26	16
Opinion	3	10	5	7	6	5	6
Genre + Task	8	9	9	6	4	3	7
Dates	11	13	9	5	7	9	9
Misc	7	19	6	8	9	2	9
Non-English	0	0	30	0	0	0	N/A
Total (Tag sample after tag-wrangling)	166	183	172	169	163	136	Total Number of Fiction Tags: 989

Nonfiction

	Always Crashing in the Same Car	The Way She Feels	Windswept	The Kissing Bug	Average
Task/Behavior	36	35	49	45	41
Genre/Subject	69	73	82	50	69
Metadata	3	5	5	13	7
Source/format	11	17	13	11	13
Motivation	16	9	9	22	14
Opinion	10	6	0	7	6
Genre + Task	5	5	8	10	7
Dates	5	6	7	7	6
Misc	12	16	4	8	10
Non-English	0	0	10	0	N/A
Total (Tag sample after tag-wrangling)	167	172	187	173	Total Number of Nonfiction Tags: 699

Other

	White Magic	Perseverance	Average
Task/Behavior	39	36	38
Genre/Subject	60	29	45
Metadata	17	31	24
Source/format	8	5	7
Motivation	9	20	15
Opinion	3	10	7
Genre + Task	3	3	3
Dates	6	17	12
Misc	2	11	7
Non-English	1	14	N/A
Total (Tag sample after tag-wrangling)	148	176	Total Number of Other Tags: 324

Excluded from dataset

	Superdoom	Something Wonderful	All the Names Given: Poems	Average
Task/Behavior	0	0	0	N/A
Genre/Subject	1	3	3	N/A
Metadata	0	1	0	N/A
Source/format	0	0	0	N/A
Motivation	0	0	0	N/A
Opinion	0	0	0	N/A
Genre + Task	0	0	0	N/A
Dates	0	0	0	N/A
Misc	0	0	0	N/A
Non-English	0	0	0	N/A
Total (Tag sample after tag-wrangling)	1	4	3	Total Tags: 8

Total number of Tags before tag-wrangling: 3,000

Total Number of Tag-sets after tag-wrangling: 2,020

Note: Genre/subject and metadata were collapsed into the ‘About book’ section and Task/behavior, Source/Format, Motivation, and Dates were all collapsed into the ‘Reader-relationship’ section. Genre + Task became the section ‘Both’ and Misc. stayed Miscellaneous. Non-English terms were excluded as researcher’s lack of multilingualism precluded them from categorizing non-English terms appropriately.