Finding (Fictional) Pandemics in the Library: Problems in the Classification of Fiction

Preston R. Salisbury
Mississippi State University, psalisbury@library.msstate.edu

Follow this and additional works at: https://digitalcommons.kennesaw.edu/seln

Part of the Library and Information Science Commons

Recommended Citation
Salisbury, Preston R. (2021) "Finding (Fictional) Pandemics in the Library: Problems in the Classification of Fiction," The Southeastern Librarian: Vol. 68 : Iss. 4 , Article 3.
Available at: https://digitalcommons.kennesaw.edu/seln/vol68/iss4/3
INTRODUCTION

As the novel coronavirus COVID-19 spread throughout the United States, it seemed reasonable to assume that people may want to search for books concerning epidemics, including fiction books and fictional epidemics as well as historical epidemics. As most libraries throughout the country closed in response to the virus, it was rendered impossible to find these by shelf browsing. This highlights the importance of subject classification of fictional works, an area that is fraught with difficulty. The purpose of this study is to highlight two problems in the subject classification of fiction which were highlighted by the COVID-19 outbreak. Using ten large public library systems throughout the Southeast, these problems are examined and identified.

It should be noted that there are different types of classification at play within fiction and literature. Subject classification, including the use of a controlled vocabulary such as Library of Congress Subject Headings (LCSH) is one type of classification. However, many libraries follow a model close to that of a bookstore and classify by genre rather than by subject. (For example, sections for general fiction, mystery, science fiction, romance, and so on.) Both classification systems are valuable for certain purposes. Patrons who know they want a certain type of book can browse and easily select a book in whatever genre is desired. One small problem with organizing by genre is that sometimes a book may fit into multiple genres. Classifying a book by subject alone suffers from a similar problem. A dedicated science fiction reader may not wish to read a romance novel regardless of the subject. It may be best, for the purpose of the cataloging of fiction, to use both LCSH or other subject headings for classification in addition to using genre terms, whether Library of Congress Genre/Form Terms (LCGFT) or other. This holds true for all forms of fiction, despite Library of Congress guidelines currently stating that topical headings are only to be used for biographical and historical fiction, and animal stories (Library of Congress, 2015).

The pandemic brought this to light more clearly, because it is impossible to browse a closed library, and while online interfaces do allow a sort of browsing, the experience is radically different. People expect to be able to find what they want quickly on the internet; while people may enjoy the browsing experience of a library or bookstore, that is not the expectation for a virtual search. Yet bibliographic records for electronic books frequently lack subject headings of any kind, and may lack genre headings as well. Many bibliographic records for electronic books have the largely useless (as far as patron discovery
is concerned) subject or genre heading of “Electronic book.” (There are various reasons why the library might wish to have this term in the system, but it does not assist with patron discovery.)

The American Library Association has only recommended providing subject headings for fiction since 1990, and, because of the sheer amount of work involved, most bibliographic records for fiction created before that date lack subject headings (DeCandido & Rogers, 1990). Doubtless, it has been difficult for many catalogers to justify the time involved in re-cataloging older materials, particularly in the public library, where newer works predominate. It is unlikely that circulation studies alone could provide significant evidence for the utility of subject headings, because if a book is famous enough to be searched for by title or author, it will circulate whether or not subject or genre terms are present within the bibliographic record. In fact, subject headings might have the most value in increasing the visibility of authors who are not as famous or who come from underrepresented groups.

**LITERATURE REVIEW**

The American Library Association first recommended a national standard practice for the “provision of genre and subject access to individual works of fiction” in 1990 (American Library Association, 2000, p. 1). The collaboration between OCLC and the Library of Congress to enhance fiction records resulted in over 15,000 records being enhanced in the ten years immediately following the project.

Research on circulation differences between fiction with subject headings and fiction without subject headings remains somewhat scanty (particularly in the Anglo-American cataloging world) and suffers from some obvious faults. One preliminary study in a single public library (McDonald, 1996) indicated no significant difference in circulation between titles with subject headings and those without. A study in an academic library (Wilson, Spillane, Cook, & Highsmith, 2000) found similar results. As of 2000, there were no studies reported in the literature finding a statistically significant correlation between bibliographic record content and use defined by circulation.

Dali and Dilevko (2006) proposed the use of additional subject headings for fiction, in addition to the usage of non-fiction or literary call numbers, to better serve the Medical Humanities (MH) in a medical center library. While this exact situation does not apply, the notion of tailoring subject and/or genre retrieval to the population served by the library certainly applies.

Saarti and Hypén (2010) discussed the formulation of a Finnish thesaurus for the purpose of fiction classification, including highlighting the difficulties associated with giving terms to fiction and the tension between specialized terms for literature studies and terms which are likely to be used by the general population looking for a book. There is a difficulty inherent in using terms that will communicate effectively to one group of users but not to another group of users, and they mentioned an additional planned thesaurus to help alleviate this difficulty.

Saarti (2019) gave an overview of differing methods of fiction classification. Several difficulties are highlighted, among them the inconsistency in description of works of fiction among various indexers and the tension between what the author terms denotative and connotative elements. Indexing the denotative elements, or the factual elements included within the work of fiction, is much simpler than indexing the connotative elements, even though these might be what the work is truly about. Traditionally, fiction indexing has stuck to the factual aspects, as well as to classifying by genre.

Mikkonen (2017) gave an intensive overview of various search strategies library patrons employ in book selection, both in the library and in the online catalog, using data gleaned from five separate studies in Finland. There were a number of findings relevant to the current study found. First, it was found that the most popular search approach was to search for a known author or title and then browse the books on the shelves, with using the library catalog being the least popular search strategy. Older participants in the study were also more likely to browse than younger participants. Interestingly, it was found that while participants used differing search strategies in enriched or traditional catalogs (enriched library catalogs allow users to interact with folksonomies in the form of tag clouds, with virtual bookshelves, and with more appealing visual features) it was possible for users of both to complete a successful search. The study also highlighted three strategies users employed in searching in library catalogs: “1. Focused querying, 2. Topical browsing and 3. Similarity based tactic” (p. 104).
Rationale

Given that multiple studies (incomplete though they may be) have failed to show a correlation between bibliographic record content and circulation statistics for fiction, why undertake a new study? The justification for it is that the previous studies largely failed to account for consumer behavior; that is, people frequently select fiction through the process of browsing rather than through a targeted searching process.

As Mikkonen (2017) observed, while known item searching is the most prominent or popular means of locating fiction within the online public access catalog (OPAC), topical searching also exists. While this study was conducted with Finnish patrons, there is little reason to suspect that American patrons employ radically different searching strategies in locating materials. Also, this study illustrates the flaw in judging the value of a subject heading by circulation statistics: since the main means of locating fiction is known item searching, the most known (and thereby most circulated) items do not need subject headings in order for patrons to locate them—at least not when they are new. However, books from smaller or independent publishers and books without much marketing (this would include many books by minority authors, particularly women of color) are not as likely to be known, and thereby not likely to be found by a known item search.

METHOD

As the COVID-19 pandemic spread, the New York Times published a list of suggested fiction to read, each of which concerned an epidemic or pandemic (Khatib, de León, Tarrg, & Alter, 2020). The list contained eleven titles (ten full-length novels and one short story.) Two titles were not used for the purpose of this study, because one was not yet released and the other so recently released that many libraries had either not yet acquired it or had not cataloged it yet. The remaining nine titles consisted of one originally published in the nineteenth century (Edgar Allan Poe’s short story The Masque of the Red Death), five published in the twenty-first century, and one of each published in the 1940s, 1960s, and 1970s. This provided a good variety of titles which span different standard practices in cataloging.

The author chose libraries for this study with a goal to be geographically representative across the southeastern United States, and selected libraries that served larger population centers within their state. Every library chosen serves a population of at least 150,000. These ten libraries include the largest population centers in Florida, Georgia, Mississippi, Louisiana, Tennessee, and North Carolina.

First, a general search was conducted for “epidemics” with the results narrowed to fiction, if possible. Libraries varied widely in terms of the capacity to narrow search by genre/form. Some OPAC or discovery layer systems were quite robust and aided discovery, while some were minimal and allowed practically no filtering of search results. Each of the result lists were examined to see which, if any, of the books appeared in the results. Next, the individual titles themselves were searched to see if the library held the titles, and the records were checked for subject headings, to see if they had any other headings related to epidemics. (It should be noted that some discovery layers employ a thesaurus that could pull up synonymous subject headings.) While some of the books did contain other relevant subject headings, they often did not add any utility to the record. For example, “plague” is a synonym of epidemic, but using the subject heading “Plagues—Fiction” on The Plague adds no discoverability beyond a known item search, unless a thesaurus is employed.

Because the goal of this part of the study was to examine how easily patrons could find works of fiction on epidemics, whatever discovery layer was readily public facing was used to conduct searches and sort results. This part of the study focused on the cataloging of the print record (or records) for these nine specific titles, although it is acknowledged that patrons could not access print titles due to the closure of the library. However, patrons could use the catalog to find a print book of interest and then separately search for an electronic version, so the quality of cataloging for the print version was still relevant.

RESULTS

As expected, nearly every library held almost every title, if only as a part of a compilation. Seven
of the nine titles were held by all the libraries; one title was held by nine of the ten libraries, and one title (Ling Ma’s Severance) was held by seven of ten libraries.

**Table 1:** Number of Titles Held by Libraries

<table>
<thead>
<tr>
<th></th>
<th>Jackson, MS</th>
<th>Memphis, TN</th>
<th>Atlanta, GA</th>
<th>Richmond, VA</th>
<th>Nashville, TN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of titles held</strong></td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td><strong>Titles found by &quot;Epidemics&quot; search</strong></td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td><strong>Titles found by &quot;Pandemics&quot; search</strong></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>New Orleans, LA</th>
<th>Miami, FL</th>
<th>Tampa, FL</th>
<th>Birmingham, AL</th>
<th>Charlotte, NC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of titles held</strong></td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td><strong>Titles found by &quot;Epidemics&quot; search</strong></td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td><strong>Titles found by &quot;Pandemics&quot; search</strong></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

However, despite most libraries holding these titles, very few of the titles came up in a general search for epidemics, narrowed to fiction where possible. On average, the general search returned 2.2 of the nine titles, with no library returning more than three or less than one. The problem becomes more apparent when looking at which books were returned by the “epidemics” search. One novel (David Koepp’s Cold Storage) was found by this search at every library; every library had used a record with the subject heading “Epidemics—Fiction”. This was also the newest of the nine books, released in September of 2019, just over six months before this study. By contrast, Michael Crichton’s The Andromeda Strain, which was released in 1969, did not appear in the general search at any of the libraries, despite being held by nine of ten libraries.

Only one library included Poe’s The Masque of the Red Death in something other than a general collection of Poe’s stories, so it is not surprising that this book did not appear in any of the searches. Of books published before 2012, Stephen King’s The Stand appeared in the general search at five libraries, likely due to the fact that this title has been re-released multiple times. Albert Camus’ The Plague appeared in the general search at a single library, likely due to the fact that that particular library held a newer re-released edition, in contrast to most libraries, which held earlier editions. Of note, although this particular library and edition did have a subject heading for epidemics, the heading was itself wrong, as it was “Epidemics—France—Fiction,” while the book is set in Algeria. While Algeria was a French colony at the time the book was written, it still would not have been considered part of France for cataloging purposes. If a geographical heading was desired, a heading for the specific city in Algeria (Oran) would be preferred to the heading for the entire country. (The bibliographic record in OCLC’s database with the most holdings does contain a geographical heading for Oran.) Another common issue with The Plague was that, while the subject heading “Plague—Fiction” is not wrong, it fails to describe the book any more than does the title, thereby providing no advantage over a known item search. Colson Whitehead’s Zone One and Justin Cronin’s Passage were cataloged using the subject heading “Virus diseases—Fiction,” which, while certainly not wrong, is unlikely to be a term used by the average searcher. This term might be used frequently within the medical community or within various scholarly communities, but it falls into the trap noted by Saarti and Hypén (2010) in that it does not fit with a large user group. But even the more commonly used term “epidemics” has fallen out of favor with many people and possibly is in need of an update.

Google Trends, which measures which terms people are searching for restricted to geography and time, reveals that Americans searched more frequently for books and movies about epidemics than about virus diseases during the time period of February 7, 2020 to May 7, 2020. In fact, there were not
enough searches for books or movies about virus diseases for Google Trends to show data for that term.

**Figure 1:** Searches for Books about Epidemics, Pandemics, or Virus Diseases

However, the term “pandemics” was used even more frequently, but this term is not yet an approved term in the Library of Congress Subject Headings. This is a case where the controlled vocabulary has not caught up with current usage. For this study, a search using the term “pandemics” at each library was also conducted, and it was found that in only one library were any of the books from the *New York*
Times list returned by a search for “Pandemics” narrowed to fiction, as seen in the table above.

Severance frequently had only genre form terms, not subject headings. Emily St. John Mandel’s Station Eleven frequently had subject headings, but not for epidemics, which is somewhat understandable because it deals more with the aftermath of the epidemic than the epidemic itself (however, two libraries did give it a subject heading of “Epidemics—Fiction”).

CONCLUSION

Despite changing practices in applying subject headings to fictional works, subject cataloging of fiction remains an inconsistent method of locating fiction. As a rule, this may not be a problem; patrons seeking fiction tend to browse the stacks or perform known item searching. However, in a situation such as the COVID-19 pandemic, browsing the stacks is impossible, and known item searching will only bring known works, decreasing the possibility of patron discovery of works which might be of interest, but which they have not heard. As noted above, this is particularly problematic for new authors, particularly authors from other cultures and authors of color.

Although subject headings are being more frequently applied to fiction than in the pre-1990 world, they often are not headings that would be used by the general public. A headline such as “Virus diseases” is useful for medical works covering viral diseases, but is less likely to be used by a patron seeking works on epidemics or pandemics, particularly works of fiction, as noted above. For nonfiction works, higher levels of granularity likely provide a benefit to users, but it is more likely that fictional works should have more general subject headings, either instead of or in addition to more specific ones. Similarly, electronic books need to be cataloged to at least the same standard as print books. While some libraries have discovery layers on their catalog that group print and electronic titles and alleviate this concern somewhat, not every library has such a system, nor can every library afford such a system.

Furthermore, as the Google Trends data shows, people are using the term “pandemic” rather than “epidemic” but “pandemic” is not an authorized term within the Library of Congress Subject Headings. There may be a variety of reasons why this problem has arisen, but it is exacerbated by the hesitance of many OPACs to allow patron tagging, which has been shown to provide important alternative access points covering areas not included within the Library of Congress Subject Headings, such as tone and emotional content (Spiteri and Pecoskie, 2016). While these tags might not fit the neutral tone and lack of bias that is a hallmark of the library profession, it is possible, when combined with editorial oversight, as Saarti and Hypén (2010) noted.

Interestingly, while the New York Times list used for this pilot study may not be particularly diverse (consisting primarily of books by white men) the standalone title that was least likely to have subject headings in the bibliographic record was by a Chinese-American woman, which could indicate that the titles most likely to need subject headings for discoverability are also less likely to receive them. This particular sample is not robust enough to draw adequate conclusions in this area, but a study of discoverability of works by non-male and non-white authors would be a useful addition to the literature in this area.

This study is far too small to draw many definite conclusions. Yet the ongoing COVID-19 pandemic has showcased, once again, the need for good cataloging even as library services move online and electronic formats become the only accessible formats. There has been insufficient study on how patrons will find and access unfamiliar works in such a situation. Also, as noted above, many (if not most) studies of the utility of subject headings for fictional works focused exclusively on circulation statistics but, as was observed, the main methods of patrons for finding fiction do not necessitate subject headings. A study of the circulation of unfamiliar works or works by new or unfamiliar authors (particularly authors of color or authors from outside the Western world) would be a useful addition to cataloging literature. Lastly, further studies and development of best practices in the area of fiction subject cataloging are needed, which would require a study of what type of subject headings are more worthwhile for patron discovery. This includes examining in fuller detail the potentials of user tagging and folksonomies for classifying fiction.
REFERENCES


APPENDIX A
Books in New York Times List and Libraries Examined

Novels in study:
Camus, Albert, The Plague
Crichton, Michael, The Andromeda Strain
Cronin, Justin, The Passage
King, Stephen, The Stand
Koepp, David, Cold Storage
Ma, Ling, Severance
Poe, Edgar Allan, The Masque of the Red Death
St. John Mandel, Emily, Station Eleven
Whitehead, Colson, Zone One

Novels also on list, but excluded from study:
Flynn, Katie, The Companions
Wright, Lawrence, The End of October

Libraries in study:
Atlanta-Fulton Public Library System, Atlanta, GA
Birmingham Public Library, Birmingham, AL
Charlotte Mecklenburg Library, Charlotte, NC
Jackson-Hinds Library System, Jackson, MS
Memphis Public Libraries, Memphis, TN
Miami-Dade Public Library System, Miami, FL
Nashville Public Library, Nashville, TN
New Orleans Public Library, New Orleans, LA
Richmond Public Library, Richmond, VA
Tampa-Hillsborough County Public Library, Tampa, FL
APPENDIX B
Sample Subject Headings for Each Novel Examined

Each of the novels covered in this study has a variety of bibliographic records which might be used, both because of multiple editions and because occasionally records are duplicated. Furthermore, libraries may edit these records (from OCLC) to suit their needs, including providing additional access points. These records are the records with the most institutional holdings attached. Subject headings and Library of Congress Genre Form Terms are as follows:

Camus, Albert, *The Plague* (OCLC # 343195)
- Epidemics—Algeria—Oran—Fiction
- Plague—Algeria—Oran—Fiction
- Psychological fiction
- Novels

Crichton, Michael, *The Andromeda Strain* (OCLC # 12231)
- Arizona—Fiction
- Extraterrestrial microorganisms—Fiction
- Thrillers (Fiction)
- Science fiction
- Novels

Cronin, Justin, *The Passage* (OCLC # 290464813)
- Vampires—Fiction
- Human experimentation in medicine—Fiction
- Virus diseases—Fiction
- United States—Fiction
- Epic fiction
- Science fiction

King, Stephen, *The Stand* (OCLC # 20490025)
- Biological warfare—Research—Fiction
- Epidemics—Fiction
- Good and evil—Fiction
- Influenza—United States—Fiction

Koepp, David, *Cold Storage* (OCLC # 1078435646)
- Epidemics—Fiction
- Science fiction
- Horror fiction
- Thrillers (Fiction)

Ma, Ling, *Severance* (OCLC # 1004911431)
- Epidemics—Fiction
- Manhattan (New York, N.Y.)—Fiction
- Black humor
- Novels

Poe, Edgar Allan, *The Masque of the Red Death* (Generally only available in compilation, OCLC # 274922 is one common example)
- Horror tales, American
- Fantasy poetry, American
- Horror fiction
- Short stories
- Poetry

St. John Mandel, Emily, *Station Eleven* (OCLC # 866615101)
- Actors—Fiction
- Action and adventure fiction
- Science fiction

Whitehead, Colson, *Zone One* (OCLC # 699763904)
- Zombies—Fiction
- Virus diseases—Fiction
- Manhattan (New York, N.Y.)—Fiction