

# Developing a Metadata Element Set for Organizing Literary Works: A Survey of the American Literary Community<sup>†</sup>

Stephen Paling

School of Library and Information Studies, University of Wisconsin-Madison, 4251 Helen C. White Hall, 600 N. Park St., Madison, WI 53706-1403 <paling@wisc.edu>



Stephen Paling is interested in the use of information technology to organize and publish information. His specific area of interest is extending Social Informatics to the study of information technology use by authors and artists. Current research efforts fall into two strands. The first involves studying the ways in which members of the American literary community use information technology, and how that use may be changing literary authorship and publishing. The second strand involves examining the ways in which the social structures embodied in literary genres can be used to organize literary works.

Paling, Stephen. *Developing a Metadata Element Set for Organizing Literary Works: A Survey of the American Literary Community*. *Knowledge Organization*, 38(3), 262-277. 25 references.

**ABSTRACT:** Various approaches have been taken to organizing literary works, but finding the most effective set of metadata elements remains an unfinished task. This paper focuses on exploring five inductively built sets for organizing new literary works for discovery by members of the American literary community. The sets feature potential metadata elements drawn from a variety of sources, including present and proposed systems, as well as prior theoretical work. The paper describes a survey study that asked members of the American literary community for input about what potential metadata elements they would be likely to use to aid the process of discovering new literary work. The paper discusses the results for each set and discusses possibilities for a new set that combines the most desirable metadata elements from each of the separate sets.

Received 30 January 2010; Revised 16 July 2010; Accepted 30 August 2010

<sup>†</sup> The author would like to thank Crystle Martin, a Ph.D. student in the School of Library and Information Studies, for conducting some of the data-gathering interviews.

## 1.0 Introduction

This paper describes a survey that is one in a series of studies meant to help establish Literature and Art Informatics, the interdisciplinary study of the design, uses, and consequences of information technologies that takes into account their role in the creative efforts of writers and artists (Paling and Nilan 2006; adapted from Kling 1999). Previous studies in the series (Paling 2008; Paling 2009; Paling and Nilan 2006) helped to establish a conceptual framework made up of four key values (Positive Regard for Symbolic Capital, Positive Regard for Autonomy, Negative Re-

gard for Immediate Financial Gain, and Positive Regard for Avant-garde-ism) that help describe the motivations of members of the American literary community in producing their work. This study differed from those previous studies, though, in focusing on the organization of literary works for discovery by members of the American literary community. The main research question in this study was, What metadata elements are likely to constitute the most effective set for organizing literary works for discovery by members of the American literary community?

In addition to addressing this general research question, this study also examined the use of one of

the key values from the conceptual framework, Positive Regard for Symbolic Capital, as a possible organizational principle for literary works. Symbolic capital as defined in earlier studies (Paling 2008; Paling 2009; Paling and Nilan 2006) includes such factors as fame and prestige, as well as appreciation of meaning and refined consumption. This definition was based, in part, on work by Bourdieu (1992) and Ross (1997). Symbolic capital can be seen in the material traces of relationships within the literary or artistic community. Writers become part of a social network that confers symbolic capital, and that network shows materially through elements such as common anthologies, magazines, and other publications that publish the work of clusters of writers. This suggests possible organizational dimensions for symbolic capital. Revealing this network through metadata may match the needs of members of the American literary community in discovering new literary work.

Metadata elements were defined as information about a literary work such as the publication history or subject content of the work. Effective was defined as useful to members of the American literary community in discovering new literary work. It was operationalized in this study as being chosen by respondents as potentially useful in deciding whether or not to read a new literary work. A set was defined for this study as a group of metadata elements. Literary works were defined as poetry, fiction, essays, or drama, with no regard to length. For example, individual poems and short stories were considered literary works, as were longer pieces of literature such as novels or volumes of poetry. Discovery was defined as the introduction of literary work by living or recently deceased authors with whose work an individual respondent was not previously familiar. No assumption was made that discovery would occur through a catalog or database search. The term was defined broadly to include hypothetical situations such as inclusion of metadata in an online literary magazine that would allow a reader to explore work related to a particular poem, short story, etc. The American literary community was defined as writers, editors, book publishers, and scholars who produce, publish, or create literary or scholarly work that appears in little magazines or literary presses that produce similar work. Little magazines were defined as described by Hoffman, Allen, and Ulrich (1947, 2): "A little magazine is a magazine designed to print artistic work which for reasons of commercial expediency is not acceptable to the money-minded periodicals or presses." The respondents carried on their literary activities in America,

were American expatriates living abroad, or worked at an organization such as a literary magazine with its principal presence in America.

This study also built on earlier, purely theoretical research (Paling 2004). That earlier research posited a classificatory horizon (596), "the material and social context within which classificatory decisions are made and in which they have efficacy in shaping discourse." That article cited similar arguments from Albrechtsen and Jacob (1998, 295), who criticized the one-size-fits-all universalism of rationalism, and empirical approaches that seek to compile large amounts of factual information about user actions. They argued further (296) that: "This implies that scheme designers are not primarily looking for ways to impose one single structure on knowledge, including one set of all-embracing facets." The approach taken in this study reflected agreement with the idea that there need not be a single structure that works for all domains. The approach in this study differed somewhat, though, in its orientation toward the gathering of empirical data. While Albrechtsen and Jacob did not rule out empirical research, this study proceeded from the explicit premise that gathering empirical evidence can form a valuable part of building knowledge organization structures for different domains. This study focused on the American literary community as one such domain.

Several other premises helped form the basis for this paper. Most specifically, this study proceeded from the premise that the question of what metadata elements will be useful to users needs to come first. We should not start from an assumption about the use of a particular delivery vehicle, for example, a MARC record and the content of its constituent fields. This study was conducted, in part, on the premise that we should consider alternatives to traditional cataloging if the tools of traditional cataloging do not meet the needs of users. It is also worth noting that the definition of literary work used in this study avoided the assumption that users will want literary works to be aggregated at the level of a book or journal. There was no a priori reason to assume that users want literary works to be aggregated at the level of books or journals, or that they want those works to be aggregated at all beyond general intellectual collocation as literary works.

## 2.0 Background

A variety of approaches to organizing literary work have been proposed or used. The study most similar to this one (Ross 2001; Ross and Chelton 2001) fo-

cuses on patrons' reasons for choosing or rejecting a book for pleasure reading. The study incorporated a purposive sample of heavy readers. That study differed significantly from this one, though, in several ways. It examined pleasure reading by patrons, not members of the literary community. It also focused primarily on reader's advisory rather than on building a set of metadata elements, and focused primarily on works of fiction.

Adkins and Bossaller (2007) examined the information about fiction offered by a variety of Web sites and OPACs, but did not solicit original user input. The study did, however, draw on previous work by Saarti (1992, 1999) and Sapp (1986). Saarti (1992, 23) described an experimental shelving arrangement for fiction that grouped books into groups such as Historical novels, Humour, and Wilderness (hunting/fishing). The library's users had largely positive reactions. The study focused on fiction, though, and did not address in detail other forms of metadata beyond shelving. Saarti (1999, 89) also examined the indexing of fiction and specifically looked at the *Finnish Thesaurus for Fiction*. Informal feedback from libraries using the thesaurus was described (91), but the study incorporated no additional responses from patrons. Sapp (1986, 495) provided an overview of different levels of subject access to fiction, but raised questions about efforts to provide subject access to fiction, and questioned whether the results are worth the effort.

The Library of Congress (LC) in America includes subject headings and classification numbers in its published tools that can describe various aspects of literary (and other) works. The American Library Association (2000) published *Guidelines on Subject Access to Individual Works of Fiction, Drama, etc.*, (GSAFD), which covered the use of *Library of Congress Subject Headings (LCSH)*. The guide discussed the syndetic structure of *LCSH*, and provided guidelines for building subdivided subject headings for literary work, primarily fiction. The *GSAFD* provided no evidence of user input into the construction of those guidelines.

The ALA SAC Subcommittee on Fiction Guidelines (Bogan et al 2003) later issued a follow-up report on the *GSAFD*. The report made several important recommendations and observations that were germane to this study. Despite having been charged with identifying a single suite of guidelines for subject and genre access to fiction, the subcommittee concluded that (Bogan et al 2003, 3): "identification of a single suite of guidelines is neither possible nor desirable at this time," adding that the "library cataloging community

must come to consensus about how to determine the subject and genre of an individual work of fiction." Further (5): "the subcommittee arrived at a consensus about the positive value of domain-specific schemes," although the *GSAFD* only refers to terms for children as one such specific domain.

It is also interesting, though, to see what the subcommittee did not do or conclude. The report referred to (Bogan et al 2003, 5): "the question of the legitimacy of the underpinnings of the [GSAFD]," and arguments "that novels don't have subjects and are not 'about' something, in the way of nonfiction works." The report did not show any evidence of serious consideration on the part of the Subcommittee of any approach other than conventional subject heading use. This study is not based on the premise that subject access should be avoided, but is based on the premise that we should not assume that traditional organizational tools such as subject headings necessarily provide the best options for organizing literary work.

Not all of the previous works focused on refinements to *LCSH* or other established thesauri. Beghtol (1994) proposed a faceted system based on theoretical principles for the classification of fiction. The proposed system, Experimental Fiction Analysis System (EFAS), was designed to produce brief classification codes somewhat similar to LC *Classification* numbers, with the difference that EFAS was intended to be a fully faceted system. EFAS comprised four major elements, Characters, Events, Spaces, and Times, as well as an Other element (Beghtol 1994, 158).

EFAS was probably most notable for avoiding the assumption that traditional cataloging standards should form the foundation for any system for organizing literary works. In that sense EFAS and the approach taken in this study bore some similarity. Other elements of EFAS bore some resemblance to traditional subject analysis, e.g., the time elements in EFAS played a similar role to the period subdivisions in *LCSH*. But EFAS differed from the approach taken in this study in several ways. EFAS, by definition, described only fiction. The current study included other kinds of literary works such as poetry and essays. The goals of EFAS centered on the needs of end-users, but the current study focused on the needs of members of the American literary community. It is possible that the metadata needs of end-user fiction readers and members of the American literary community will overlap, but this study did not address that question.

The prior works on the organization of literary work largely ignored poetry, except for Chan (2005). Chan's work mentioned genre headings for poetry,

e.g., Poetry (287), as well as the use of generic subdivisions related to poetry, e.g., Nature—Poetry (294). The focus was largely on the rules for assembling LC subject headings.

OCLC's FictionFinder (<http://fictionfinder.oclc.org/>) is a notable project from the bibliographic community. It provides metadata about fiction works included in WorldCat (<http://www.worldcat.org/>), and provides metadata elements such as the number of published editions of a work and the number of libraries that hold the work. It is a FRBR-based prototype that uses WorldCat records (OCLC 2010). FRBR was developed as "an entity-relationship model as a generalized view of the bibliographic universe" (Tillett 2003, 2). FRBR attempts to take into account factors such as part-whole relationships and different manifestations of the same work (Tillett 2003, 2, 5). FictionFinder was on hiatus at the time this paper was written, so the metadata elements it offers may change.

The inclusion of metadata elements such as the number of editions of a book also touches on the debate over the definition of a *work* for bibliographic purposes. For example, Yee (1994, 19) argued for the importance of alerting users to the existence of multiple editions of a work. That work provides interesting context to the choice to provide that particular piece of metadata in FictionFinder.

A variety of other databases provide similar metadata or the full text of literature from various periods, nations, and groups. Many of the databases focus on access to older works rather than currently publishing authors. None of these tools is tailored specifically to members of the American literary community, whose needs may not be the same as library patrons who are not themselves members of the community. Appendix A includes URLs for a number of those resources.

### 3.0 Methodology

This study examined five inductively built sets of potential metadata elements for the organization of literary works that were drawn both from traditional cataloging, and from other sources such as online booksellers. The study instrument also allowed respondents to suggest potential metadata elements not included in the inductively built sets. The five sets comprised 43 elements (42 elements plus an Other category). Appendix B lists the individual elements by set.

#### 3.1 Bibliographic Element Set

These are elements that one would expect to find in standard bibliographic records (e.g., classification numbers and *LCSH* subject headings), as well as less-traditional uses of bibliographic records. For example, as mentioned above, OCLC's FictionFinder gives prominent placement to metadata elements such as the number of libraries that hold a book and the number of editions of a book.

#### 3.2 Faceted Element Set

This set comprised metadata elements found in, or based on, Beghtol's (1994) EFAS system. In addition to relatively common metadata elements such as the time period in which a piece of fiction is set, the EFAS system included more unusual metadata elements such as relationships between characters, e.g., mother-daughter, and the names of known fictional characters.

#### 3.3 Popular Element Set

This set comprised metadata elements that one might find in popular sources such as online booksellers, including elements such as reader/user comments. This also included metadata elements such as a sample of a work's content, which is a common feature of online bookseller sites.

#### 3.4 Shared Element Set

This set comprised metadata elements that did not belong clearly in any of the other sets because the elements are so common, e.g., the name of the author(-s), the length of the work, and the year in which a work was published.

#### 3.5 Symbolic Capital Set

This set comprised metadata elements based on a work's place in literary writing as a genre, e.g., the names of awards given to authors or works, or the names of authors published in the same literary magazine. This set of metadata elements was built on previous work done in this series of studies (Paling 2008; Paling 2009; Paling and Nilan 2006). An attempt was made to produce sets of elements that together would provide a broad and balanced selection for respondents to evaluate. However, the Other category was included in order to allow for unanticipated elements suggested by the respondents.

The goal was not to choose one of the metadata element sets over the others and use that set exclusively. These inductively built sets served primarily to provide an approximate indication of how well a number of basic approaches are suited to the needs of the American literary community. The larger goal is to build a set of metadata elements, probably drawn from all of the smaller sets as well as from respondent input, that will make up a new set of metadata elements for organizing literary work for discovery by members of the that community.

The difficulty in assigning some of the potential elements to a set provided one of the reasons for not attempting to choose a single set. For example, one of the potential metadata elements was the name of a work's author. That element became part of the Shared metadata element set because its use is so common in so many contexts, including library catalogs, literary databases, and literary Web sites. It would be unusual to not see the author's name. But an argument could have been made that the author's name belonged in the Symbolic Capital metadata element set, since an author's name may come to the reader freighted with the author's reputation, or lack thereof. Other elements, e.g., library-assigned subject headings, clearly belonged in a specific set. Taken as a group, the element sets provided a rough way of looking at how the respondents viewed various basic approaches, which offered evidence about what a unified, inductively built set of metadata elements might look like.

### 3.6 Research Questions

In order to address the main research question (What metadata elements are likely to constitute the most effective set for organizing literary works for discovery by members of the American literary community?), the study addressed two smaller research questions:

RQ1: What means do members of the American literary community currently use to discover new literary works?

RQ2: Which metadata elements do members of the American literary community find potentially useful for discovering new literary works?

RQ1 provided important background information about the current general means that members of the American literary community use to discover literary works. Knowing, for example, that community mem-

bers value word-of-mouth recommendations highly might suggest that symbolic capital in the form of membership in social networks plays a significant role in literature discovery. That, in turn, might suggest that metadata elements that take advantage of a work's placement in a social network might be useful to community members. RQ2 provided the principal data for the study. Knowing which potential metadata elements respondents prefer will take us a significant step toward building a suitable set for organizing literary work.

### 3.7 Sample

The sampling frame comprised membership lists from the Modern Language Association (MLA), The Association of Writers and Writing Programs (AWP), as well as information from *Poet's Market*, *Novel & Short Story Writer's Market*, and the CLMP *Literary Press and Magazine Directory*. No central membership list or organization exists for the American literary community. However, these lists combined provided access to contact information for writers, magazine editors, book publishers, and literary scholars. Because the lists reflected institutional affiliations in many cases, the sampling frame tended to contain members of the more established parts of the field. The lists were pooled, duplicates were removed when possible, and a random sample was drawn for contact.

Most of the invitations were sent by e-mail, but some respondents were also contacted by phone to determine whether respondents who did not have a publicly available e-mail address showed a different pattern of responses than respondents who did have a publicly available e-mail address. No invitations were sent by postal mail. In addition, some of the questions were assumed to be in a formative state, and possibly in need of refinement. Because of this, the interviews were conducted by phone so that informal, spontaneous feedback about the questions was possible. A postal mail invitation for a phone interview would have constituted a very cumbersome method for arranging interview times, so no postal mail invitations were sent. Future studies may involve mailed questionnaires. One respondent was given a written questionnaire to fill out because of an impairment that made a phone conversation impractical.

### 3.8 Interview Instrument

Along with the principal questions about metadata elements, the interview instrument included questions

about each respondent's background. In addition to basic demographic facts such as age and gender, the background information included each respondent's professional situation, the type of writing each did, and what types of publications each had to her or his credit.

The survey instrument was fairly long, and this paper only addresses a subset of the questions. The questions not reported on here were discussed in a separate paper (Paling and Martin 2009). Those questions dealt with general technology use by members of the American literary community, and looked at whether and how information technology use is changing the basic values of that community. Other than the general background questions, there was little overlap between the questions addressed in that paper and the questions addressed here. The exception to that is the use in both papers of the idea of Symbolic Capital. Positive Regard for Symbolic Capital is one of the key values addressed in the other paper, and it is tested here as a potential organizational principle. The instrument contained 23 basic questions, but several of the questions included multiple items that could be viewed as questions themselves. The total number of items was 102. The focus in this paper was on the 43 items that were potential metadata elements, as well as the questions that covered basic background information.

The first block of metadata-related questions asked respondents about their current sources for discovering new literary work (RQ1). The initial question for that section was, "For each of the following sources, please answer Yes or No whether you use, or have used, that source to discover new literary work." This list of sources was relatively short:

1. A database of literary work, e.g., Fiction Finder or LitFinder.
2. A library catalog.
3. Book reviews.
4. Literary magazines.
5. Literary readings or similar live events.
6. Multiple-author anthologies, e.g., the annual Pushcart Prize anthology.
7. Publishers' print catalogs.
8. Publishers' Web sites.
9. Scholarly articles.
10. Scholarly books.
11. Word of mouth.
12. Are there any other sources that I have not named on this list?

The first 11 items were rotated randomly to prevent any order effect in how the respondents answered. Because of the length of the instrument more precise responses were not requested. Follow-up studies may ask for more detailed responses such as importance, frequency of use, etc.

The second, and principal, block of metadata-related questions was longer, and asked respondents to indicate whether they would use the items on the list (See Appendix B) in deciding to read a new piece of literary work (RQ2). These items were also rotated randomly except for an Other category. The initial question for that section was:

For the next question I would like to ask you about characteristics of new literary work. For each of the items, please indicate with Yes or No whether you would use that characteristic in deciding whether to read a piece of new literary work. For example, if you would use the presence of cover art on a new book when you decided whether to read the book, you could answer Yes.

The question included the phrase "deciding whether to read a piece of new literary work" in part to avoid technical jargon related to the organization of information and any preconceptions the respondents might have about that jargon. The decision about whether or not to read a piece of literary work is central to the process of discovery that is part of the focus of this study. This study sought to begin finding out what potential metadata elements would be most useful to members of the American literary community in making that decision. Cover art was used as an example because it was not itself on the list of potential metadata elements, but nonetheless provided a similar example. Using an actual item from the list might have created biased responses about that item.

As with the first block of questions, follow-up studies may ask for more detailed answers. But a significant goal for this study was to narrow the range of potential metadata elements. The resulting list can be refined in subsequent studies. The data from the questions were analyzed primarily using descriptive statistics. There was no a priori way to establish any particular cutoff above or below which we can say that the item should automatically be included or excluded in an eventual metadata scheme.

## 4.0 Results

### 4.1 Response Rate and General Information

A total of 237 invitations were sent out, all but 12 by e-mail. One of those potential respondents was deceased, and two messages bounced, leaving 234 viable invitations, with 93 respondents agreeing to participate (a response rate of 39%). Nine people who initially agreed to participate did not, leaving a final response rate of 36% from the viable invitations. This resulted in a sample of 84 respondents. A total of 12 potential respondents were contacted by phone. The respondents contacted by phone were part of the random sample, but no e-mail contact information could be found for them. One of those potential respondents was deceased, leaving 11 viable invitations, with 5 respondents agreeing to participate (a response rate of 45%). The sample included 43 women and 41 men. The median age of the respondents was 52. The median age for the United States as a whole is 36.8 (U.S. Census Bureau 2008). Two respondents declined to provide their age.

Some question attrition occurred during the interviews. Out of the total pool of 84 respondents, 78 (93%) answered all of the questions. The instrument was fairly lengthy, and contained 102 total items, of which this paper covers only a subset. The total of 102 includes all of the individual items about which the respondents were asked. The actual number of questions was 23. With a sample of 84, and 102 items on the instrument, the total number of items for the entire survey was 8,568. The total number of non-responses was 62 (<1%), with 35 (56%) of those non-responses coming from a single respondent who did not finish the interview. This was a significant number of questions for respondents to answer, but that was mitigated by the brevity of the individual items, most of which called for simple answers, often a single word. The random ordering of many of the items also helped insure that any fatigue effects would be distributed evenly across the questions.

Table 1 shows a breakdown of the professional situations of the respondents (rounding error causes a total percentage of 101% to appear). A plurality of the respondents was made up of literary writers, with the next-largest group being literary scholars. The respondents were allowed to choose whatever category each felt was most appropriate. For example, some of the respondents who defined themselves as Literary Writers make their living as academics, and some do not. So the choice of professional categories should

be taken as representative of how each respondent perceived his or her primary role, not as a dispositive categorization with no overlap. Many members of the literary community have multiple roles, but nonetheless feel the strongest affinity for one of those roles.

Professional Situation	Percentage
Literary Writer	40
Literary Scholar	20
Other	11
Book Editor	10
Mixed	10
Magazine Editor	10

Table 1. Respondent Percentage by Professional Role.

The respondents who chose Other described a variety of professional situations, including high school teachers, administrators of arts or academic programs, and academics or graduate students in other disciplines. No one description dominated within Other.

All but 2 of the respondents had had work published. Over half of the respondents (48 out of 84, 57%) had had a book published. A similar number (51 out of 84, 61%) had been anthologized. Just under half had had a scholarly article published (40 out of 84, 48%), and almost three quarters (60 out of 84, 71%) had had an individual work published in a literary magazine or similar publication. The respondents were allowed to choose more than one category of publication.

### 4.2 RQ1: What means do members of the American literary community currently use to discover new literary works?

A majority of the respondents used each category of sources except for literary databases. It emerged anecdotally in the interviews that many of the respondents had not heard of databases of literary work. While this might seem at first to be a rejection of those tools, it also suggests that databases of literary work might gain users with increasing exposure and awareness. Figure 1 shows the results. So, for example, 96% of the respondents indicated that they used word of mouth as a way to discover new literary work.

The Other category comprised a number of responses. Some of the responses were actually redundant with the categories provided. For example, several respondents mentioned that they considered

## Sources of New Literature

n=84, multiple choices possible

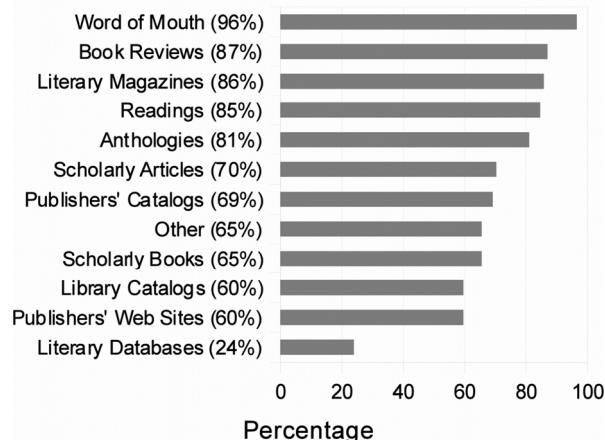


Figure 1: Sources of New Literature Used by Respondents

such technological tools as listservs and use of Google to be an extension of Word of Mouth. A response like that one addressed the medium through which word-of-mouth comments are delivered, but not the source (a colleague or other commenter). Other sources that were mentioned included browsing in bookstores (both online and off), submissions to a magazine at which the respondent worked, and agents. No single response from the Other category garnered more than a relative handful of responses. Some of the sources, e.g., writers' blogs, may be included in future instruments, but at present a breakout of those sources from Other would not change the basic shape of the upper parts of the distribution. Those responses would simply add a somewhat longer, lower tail to the distribution.

Table 2 shows the results for sources of new literary work separated by contact method. The cell values represent the percentage of Yes responses for each source. For example, 96% of the respondents contacted by e-mail indicated that they use word of mouth to discover new literary works.

Some of the differences between the groups are small, but one pattern is very clear. For 9 of the 12 categories, the respondents contacted by phone showed a higher percentage of Yes responses. This suggested two things. First, the respondents contacted by phone seemed to be at least slightly more frequent users of most of the sources. Because the availability of e-mail addresses may have reflected underlying differences in technology use, this result couldn't be ignored. One mitigating factor, though, was the fact the both groups showed the same basic lack of pattern in the Other category. In other words,

Source	E-mail	Phone
Word of Mouth	96	100
Book Reviews	86	100
Literary Magazines	85	100
Readings	85	80
Anthologies	80	100
Scholarly Articles	70	80
Publishers' Catalogs	67	100
Other	65	80
Scholarly Books	66	60
Library Catalogs	58	80
Publishers' Web Sites	58	80
Literary Databases	24	20

Table 2. Sources for New Literary Work (Percent Yes) by Contact Method.

both groups used the same range of sources, but with different frequency.

### 4.3 RQ2: Which metadata elements do members of the American literary community find potentially helpful for discovering new literary works?

The potential metadata elements drew widely varying levels of Yes responses. The potential metadata element that garnered the most interest (the name of the work's author) was of potential interest to 93% of the respondents. The potential metadata element that garnered the least interest (the number of libraries holding the book or journal in which a work appeared) was of potential interest to only 6% of the respondents. The Other category was chosen by 51% of the respondents. The following tables (3 through 7) show the percentage of Yes responses for each set, with each table containing the potential items for one of the sets.

Potential Bibliographic Element	%Y
Information about earlier works on which a work is based, e.g., folklore on which a current story is based.	80
Information about derivative works based on the work in question, e.g., a children's version of a novel.	45
A library-assigned subject heading.	31
The number of editions of a book that have been published.	20

Potential Bibliographic Element	%Y
A library classification number such as a Dewey Decimal number or Library of Congress classification.	14
The names of libraries which hold the journal or book in which the work appeared.	11
The number of libraries which hold the journal or book in which the work appeared.	6

Table 3. Results by Potential Bibliographic Element.

Potential Faceted Element	%Y
Information about the era depicted in a literary work, e.g., the period or age in which actions take place, or the time of day described in the work.	71
Information about the geographic locations in a literary work, e.g., locations of characters or events in a fictional world.	64
Information about events in a literary work, e.g., physical actions by characters, or larger phenomena such as warfare.	63
Factual information available in the work, e.g., the actions of historical figures.	62
Information about characters in a literary work, e.g., a character's religion or occupation.	49
Relationships between characters, e.g., mother-daughter.	45
The names of known fictional characters.	39

Table 4. Results by Potential Faceted Element.

Potential Popular Element	%Y
A sample of the work's content.	89
The specific genre into which a work falls, e.g., mystery or adventure.	85
Comments from other readers.	74
An indication of the author's intent for the work, e.g., humor.	67
A description of the theme of the work, e.g., Nature or Truth.	64
The intended audience for a work, e.g., popular vs. literary.	56
An indication of the intended emotional experience of the work, e.g., tearjerker.	46

Table 5. Results by Potential Popular Element.

Potential Shared Element	%Y
The name of the author.	93
The language in which the work is written.	92
The broad genre into which a work falls, e.g., poetry or fiction.	88
The broad form which a work takes, e.g., short story, poem, collection, or novel.	82
An abstract or similar content description.	82
The title of the work.	70
The age of the intended audience.	56
The author's nationality or ethnic background.	54
The length of the work.	42
The year in which a work was published.	37

Table 6. Results by Potential Shared Element.

Potential Symbolic Capital Element	%Y
The name of an award given to the author.	81
The name of an award given to the work.	76
The title of a literary magazine in which the work has appeared.	75
The title of an anthology in which the work has appeared.	67
A list of authors whose work appeared in the same anthology.	67
The name of the publisher of the book or magazine in which the work appeared.	63
A list of authors who have also had books published by a particular book publisher.	61
A list of authors whose work appeared in the same issue of a literary magazine.	60
A list of authors whose work appeared in previous issues of the same literary magazine.	55
A list of authors whose work appeared in a previous edition of the same anthology.	54
The name of the editor or editors who decided to publish the work.	51

Table 7. Results by Potential Symbolic Capital Element.

Responses in the Other category described a mix of unanticipated possibilities, as well as a number of responses that were largely redundant with the inductively built list of categories. For example, a number of respondents reiterated the importance of word-of-mouth recommendations, in some cases emphasizing that those recommendations had the most value

	Shared	Symbolic Capital	Bibliographic	Faceted	Popular	Other
E-mail	70	64	30	56	68	51
Phone	76	67	26	54	77	40

Table 8. Potential Metadata Elements (Percent Yes) by Set.

when they came from colleagues. Similarly, a number of respondents discussed such elements as the “writing itself” and the “first few pages” of a work, both of which were largely redundant with “A sample of the work’s content,” which was one of the elements of the Popular set. The Other responses that did not match any of the elements from the inductively built sets were largely scattered responses, e.g., availability in large print, that occurred once or a very small number of times. The only such response that was repeated described various forms of matching the respondent’s personal interest in some subject area.

Table 8 shows the results by set separated by contact method. The Yes responses were pooled for each set, and calculated as a percentage of the total responses for that set. So, for example, the respondents contacted by e-mail gave 70% Yes responses for items in the Shared set. Pooling the responses provided a rough view of how the members of both groups viewed each. The differences in the percentage of Yes responses were smaller and more evenly distributed than for the sources of new literary work in RQ1. Including the Other category, there were six groups of pooled answers. For three of the six the respondents contacted by e-mail gave more Yes responses, and for the other three groups the respondents contacted by e-mail gave more Yes responses.

Figure 2 contains a radial polygon that is based on the percent Yes response for each metadata element set as a percentage of the total responses in each category. The percentage for each set is included in the graph. So, for example, 70% of the responses to potential elements in the Shared set were Yes. If all five of the sets had drawn an equal percentage of Yes responses, the pentagon in the graph would be symmetrical. That is largely the case except for the Bibliographic set, which drew only 30% Yes responses. That was the only set that, as a whole, fell below 50%.

There is no simple way to determine what percentage of Yes responses might serve as a cutoff, i.e., a point below which a potential metadata element should be excluded from use. But at least some indication might be provided by looking at the shape of the distribution. Figure 3 shows the shape of the distribution, with the items ranked from highest percentage of Yes responses to least. Because the focus in this instance is on the general shape of the distribution, Figure 3 does not include labels for the individual items.

The shape shows a clearly steeper downward trend for the last few potential elements. Those potential elements were all drawn from the Bibliographic set. Table 7 shows the percentage of Yes responses for each of the top 10 and bottom 10 items, as well as the difference ( $\Delta$ ) from the previous element’s percent-

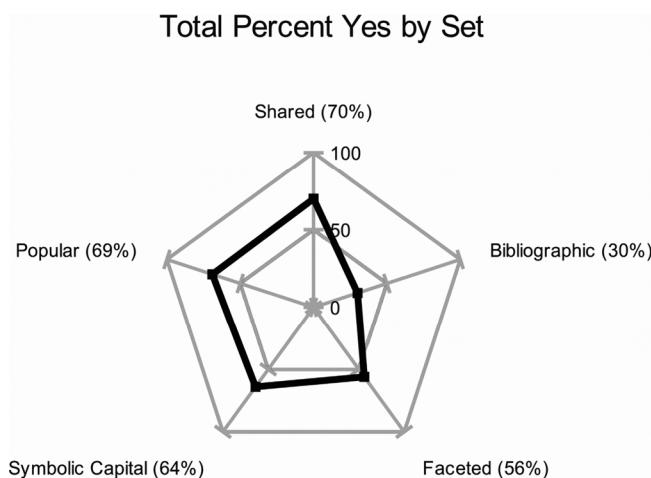


Figure 2. Percent Yes Responses by Set.

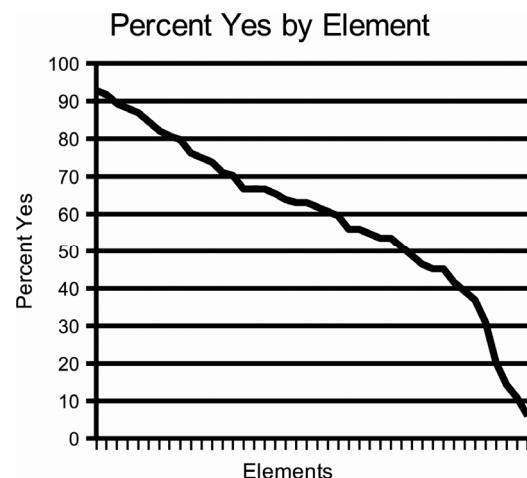


Figure 3. Percent Yes by Ranked Element.

	#	1	2	3	4	5	6	7	8	9	10
Top 10	%	93	92	89	88	87	85	82	81	80	76
	Δ	NA	-1	-3	-1	-1	-2	-3	-1	-1	-4
Bottom 10	%	45	45	42	39	37	31	20	14	11	6
	Δ	-1	0	-3	-3	-2	-6	-11	-6	-3	-5

Table 9. Percent Yes and Decline from Previous Element, by Top 10 and Bottom 10 Percent Yes Responses.

age. For example, item 2 in the top 10 drew 92% Yes responses, a decline of 1% from the previous element. The Δ does not apply to element 1 in the top 10 because there was, by definition, no preceding element. The top 5 elements, in order, were: Author Name, Language of the Work, Sample of the Work, Broad Genre of the Work, the Broad Form of the Work. All of the top 5 elements were drawn from the Shared or Popular element sets.

The steeper decline starts with items 6-10 in the bottom 10. These elements (in order: Subject Headings, Number of Editions, Classification Number, Names of Libraries Holding a Work, Number of Libraries Holding a Work) were all drawn from the Bibliographic set. Because of the relatively low percentages and accelerated decline, this might be a plausible place to cut the list before additional testing of potential metadata elements. Another plausible point might be to cut at 30% Yes responses because of the extremely steep drop after that point.

## 5.0 Conclusions

### 5.1 RQ1: What means do members of the American literary community currently use to discover new literary works?

The results suggest that many members of the American literary community are omnivores in choosing their means for discovering new literary work. No one of the potential sources dominates. A majority of the respondents use each of the potential means for discovering new literary works, with the lone exception being databases of literary work. It emerged anecdotally during the interviews that many of the respondents do not know of the existence of such databases. This suggests that room exists for a growth in use of those databases, and the findings of this study could help shape the metadata elements included in such databases.

It is interesting to note that 3 of the top 5 choices (Book Reviews, Literary Magazines, and Antholo-

gies) involve some level of direct editorial vetting. Anecdotally, many of the respondents also made it clear that Word of Mouth (with the highest percentage of Yes responses) for them referred to discussions with colleagues and friends whose judgment those respondents trusted, which is itself a form of vetting. Many readings (with the fourth highest percentage of Yes responses) also involve vetting and invitations, but those dimensions were not captured in the survey instrument. Overall this selection of sources would tend to support the finding of the importance of symbolic capital in earlier studies (Paling 2008; Paling 2009; Paling and Nilan 2006). While symbolic capital does not predominate, it clearly plays a role in how members of the American literary community decide what to read. Vetting partly defines the idea of symbolic capital as explained by Bourdieu (1992), who argued that the symbolic capital of a work's discoverer is inscribed "in the relationship with the writers and the artists he or she supports ('a publisher', says one of them, 'is his catalogue,...') (168). So the vetting that occurs in these relationships plays a key role in defining symbolic capital.

### 5.2 RQ2: Which metadata elements do members of the American literary community find potentially useful for discovering new literary works?

Members of the American literary community view a variety of metadata elements as potentially useful for the discovery of new literary work. No one of the five sets used in the study predominates, but the Bibliographic set does draw less interest than the other sets. This validates the premise that we should consider changes or alternatives to traditional cataloguing practice.

We also need to consider how widely applicable some of the potential metadata elements will turn out to be. For example, the inclusion of a historical figure or established literary character may not be a factor in many literary works. In contrast, any published work, by definition, has at least a brief publication history.

A low rate of Yes responses and/or narrow applicability of some potential elements may serve as the best guide for cutting down the list of metadata elements to those elements most likely to be useful to members of the American literary community.

The elements from the Symbolic Capital set were the only set of elements in which each potential element garnered Yes responses from a majority of the respondents. These elements, while not predominant, can add useful dimensions to metadata about literary works. For example, the publication history of a poem or short story that has been published in a literary magazine, anthologized, and included in a book by the author, could provide a useful description of where the work fits in the larger literary community. The picture that begins to emerge is of a literary work embedded in a social network, rather than a network of subject terms or classification numbers. No system for organizing literary works currently includes that information on a consistent basis. The elements from the Symbolic Capital set should be considered strong possibilities for inclusion in any system for organizing literary works for discovery by members of the American literary community.

It is interesting to note that it is implicit in these results that the respondents do not necessarily want literary works aggregated at the level of a book or magazine. For example, one of the questions asked whether each respondent would consider the broad form of a work (poem, novel, etc.) when deciding whether to read the work. That item drew 82% Yes responses and the example given of an individual poem suggests a desire for individual works, not just aggregated works. Similarly, when asked whether they would be interested in a list of authors published in the same issue of a literary magazine, 60% gave Yes as a response. In other words, the respondents didn't want to know just the magazine title. They also wanted to know at least some information about authors of the other individual works in that magazine.

### 5.3 Overall

No one element or set predominates, although the Symbolic Capital set is notable for both for having all of its items garner interest from a majority of the respondents. The Bibliographic set did stand out negatively. The other four sets all drew fairly strong interest from the respondents. The decline in the percentage of Yes responses from the top of the ranked list of potential metadata elements to the bottom is gradual and essentially linear through nearly the bottom

of the list, at which point the percentage of Yes responses drops sharply for a cluster of potential metadata elements from the Bibliographic set. This suggests that we should consider using a variety of metadata elements to aid the discovery of new literary works by members of the American literary community, and a mixed element set is most likely to meet the needs of community members. We will need more research, though, to determine what mix of metadata elements will constitute the most effective set.

Although the findings here are preliminary, it is worth pointing to some of the possibilities implicit in these results. For example, the success of the Symbolic Capital set suggests that the respondents see literary works as, at least in part, embedded in a social network of editors, publishers and authors, with publications serving as the medium for that network. This would be particularly interesting in a Web-based environment where links can readily be made between, for example, a poem in an online literary magazine and a copy of that poem that appears in an online anthology. That would allow readers to explore work by other authors in that anthology without having to use another application or database. Some of that information appears in various current metadata records, e.g., the publisher's name. But in many cases the connection between a poem in a magazine and that same poem in an anthology might require multiple searches across multiple sources.

One intriguing finding is the evidence that many members of the American literary community remain unaware of the presence of databases of literary work such as OCLC's FictionFinder. On one hand, it suggests that such databases have a small audience within the community. But on the other hand, it suggests the possibility for the use of databases of literary work to grow. It is possible that this line of research might lead to adjustments that will bring databases and portals of literary work closer to the metadata element set that members of the American literary community seem to want.

We also need to consider whether some of the elements that the members of the literary community want are actually practical to deliver. For example, we know that a work's publication history is often easy to obtain from the author, and in fact, may be required in some circumstances such as publication of a book that includes previously published poems. In contrast, it is much less clear whether authors and editors will be willing to contribute information such as a description of the intended emotional experience of their work, or whether any clear basis exists on

which a cataloger or indexer could form such a description. In a similar vein, it is not clear whether catalogers will have the time needed to provide subject access on the level of individual poems or stories inside an anthology or monograph.

Some debate is possible about the assignment of elements to the different sets. But elements such as library-assigned subject headings and classification numbers are indisputably at the heart of contemporary cataloging practice. Those elements draw low levels of interest from members of the American literary community. While elements such as classification numbers may be useful for retrieval of a known item, they do not appear to play a significant role in the discovery of new literary work by community members. The overall results suggest that we should consider building a new, hybrid set for the discovery of new literary work by members of the American literary community.

## 6.0 Limitations and Future Research

The relatively small sample represents the primary limitation of this study. Only a small number of the items on the instrument need further development, but it was not possible to determine that at the outset. So the sample size made sense for a study that was larger than a pilot study, but small enough to reduce the risk of investing significantly greater effort in an instrument that might have produced more problems than actually turned out to be the case.

In a related vein, the sampling frame also provided limitations. The sampling frame included predominantly members of the more established portions of the American literary community, and so may have been less effective at capturing the opinions of less-established members of the community. That particular limitation will be very difficult to address fully. Because membership in the American literary community is not clearly defined, and because less-established members virtually by definition appear less frequently in directories, mailing lists, and other sources, it may never be possible to establish a sampling frame that includes all members of the community. Less-established members may need to be reached through purposive sampling based, for example, on occasional publication in little magazines. This kind of sampling was used in the previous studies. Taking these limits into account, though, we can still reach moderately strong conclusions about the more-established portions of the American literary community.

This study focused on metadata elements of direct interest to users. Metadata elements such as the names of libraries that hold a particular work may be of professional interest for purposes such as ILL, but the respondents themselves showed little interest in such information.

The methodological evidence about the contact methods is somewhat ambiguous. Contact by phone resulted in a higher response rate, but not dramatically so. Similarly, the respondents contacted by e-mail and the respondents contacted by phone showed mildly different patterns in their use of sources for new literary work and their preferences for potential metadata elements. The differences were mainly differences in proportion, though, and not actual differences in the range of sources or potential metadata elements. E-mail contact seems to be sufficient for developmental studies, but contact by postal mail might be suitable for later studies that might involve a briefer instrument containing fully developed items. Contact by postal mail would provide the advantage of using a contact method that is unlikely to be a confounding technological variable, as postal addresses are available for virtually all members of the sampling frame. The preliminary evidence suggests, though, that this is not a vital issue.

In hindsight the focus on literary works that are new, and not just new to the respondent, created an unnecessary limit on the results. Future iterations will cover literary works that are new to the respondent, but not necessarily new more generally.

Another clear limitation stems from the study's focus on the American literary community. Rather than viewing this only as a limit, though, we can also view the focus of this study as a point from which this research agenda can continue. For example, at least two other paths of comparative research present themselves: community and content. Future studies could very productively address differences across different cultural groups, as well as different art forms. Some of the findings may be common to a variety of cultural groups and across multiple art forms, while other findings may change along both of those vectors. So the apparent limit points to a potential long-term research agenda.

## References

Adkins, Denice and Bossaller, Jenny E. 2007. Fiction access points across computer-mediated book information sources: a comparison of online bookstores, reader advisory databases, and public library

catalogs. *Library & information science research* 29: 354-68.

Albrechtsen, Hanne and Jacob, Elin K. 1998. The dynamics of classification systems as boundary objects for cooperation in the electronic library. *Library trends* 47n2: 293-312.

American Library Association. 2000. *Guidelines on subject access to individual works of fiction, drama, etc.* 2nd ed. Chicago: American Library Association.

Beghtol, Clare 1994. *The classification of fiction: the development of a system based theoretical principles.* Metuchen, NJ: Scarecrow.

Bogart, Ruth, DeSio, Sandra M., Gabel, Linda, Gilchrist, Jane, Hearn, Stephen, Hoffman, Shannon, et al. (2003). *Final report of SAC Subcommittee on Fiction Guidelines.* Chicago: American Library Association. Retrieved January 12 2010, from <http://www.ala.org/ala/mgrps/divs/alcts/mgrps/ccs/cmtes/sac/inact/fictionguideline/SACFicGuideFinalRpt.doc>.

Bourdieu, Pierre. 1992. *The rules of art: genesis and structure of the literary field.* Stanford, CA: Stanford University.

Chan, Lois Mai. 2005. *Library of Congress subject headings: Principles and applications* 4th ed. Westport, Conn.: Libraries Unlimited.

Greenberg, Jane. 2004. User comprehension and searching with information retrieval thesauri. *Cataloging & classification quarterly* 37n3/4: 103-20.

Hoffman, Frederick J., Allen, Charles, and Ulrich, Carolyn F. (1947). *The little magazine: a history and bibliography.* Princeton: Princeton University.

Kling, Rob. 1999. What is Social Informatics and why does it matter? *D-Lib* 5. Retrieved September 12, 2006, <http://www.dlib.org/dlib/january99/kling/01kling.html>.

OCNL. 2010. *FictionFinder is on hiatus.* Retrieved July 7, 2010, from <http://fictionfinder.ocnl.org/>.

Paling, Stephen. 2004. Classification, rhetoric, and the classificatory horizon. *Library trends* 52n3: 588-603.

Paling, Stephen. 2008. Technology, genres, and value change: literary authors and aesthetic use of information technology. *Journal of the American Society for Information Science and Technology* 59: 1238-51.

Paling, Stephen. 2009. Mapping techno-literary spaces: adapting multiple correspondence analysis for literature and art informatics. *Information research* 14v2. Retrieved August 3, 2009, from <http://informationr.net/ir/14-2/paper401.html>.

Paling, Stephen and Nilan, Michael. 2006. Technology, values, and genre change: the case of little magazines. *Journal of the American Society for Information Science and Technology* 57: 862-72.

Paling, Stephen and Martin, Crystle. 2009. *Transformative use of information technology in american literary writing: A survey of literary community members.* Paper presented at the Social Informatics Symposium at the Annual Meeting of the American Society for Information Science and Technology, Vancouver, BC.

Ross, Catherine S. 2001. Making choices: what readers have to say about choosing books to read for pleasure. *The acquisitions librarian* 13n25: 5-21.

Ross, Catherine S., and Chelton, Mary K. 2001. Reader's advisory: matching mood and material. *Library journal* 126n2: 52-55.

Ross, Trevor. 1997. "Pure poetry": cultural capital and the rejection of classicism. *Modern language quarterly* 58: 437-56.

Saarti, Jarmo. 1992. Experiments with categorizing fiction in Lohtaja Library. *Scandinavian public library quarterly* 25n4: 22-24.

Saarti, Jarmo. 1999. Fiction indexing and the development of fiction thesauri. *Journal of librarianship and information science* 31n2: 85-92.

Sapp, Gregg. 1986. The levels of access: subject approaches to fiction. *RQ* 25: 489-97.

Tillett, Barbara B. 2003. *What is FRBR? A conceptual model for the bibliographic universe.* Washington, D.C.: Library of Congress.

U.S. Census Bureau. 2008. *Population estimates.* Retrieved November 29, 2009, from <http://www.census.gov/popest/national/asrh/NC-EST2008/NC-EST2008-01.xls>.

Yee, Martha M. 1994. What is a work? Part 1: The user and the objects of the catalog. *Cataloging & classification quarterly* 19n1: 9-28.

## Appendix A: Selected Online Sources of Literary Works

Booklist Online:

<http://www.booklistonline.com/>.

Bowker's Fiction Connection:

<http://www.fictionconnection.com/>.

EBSCO's NoveList:

<http://www.ebscohost.com/novelist/>.

Fictiondb:

<http://www.fictiondb.com/>.

Gale's LitFinder:

[http://www.gale.cengage.com/LitSolutions/lit\\_resources/litfinder/](http://www.gale.cengage.com/LitSolutions/lit_resources/litfinder/).

Gale's Books and Authors:

<http://www.gale.cengage.com/booksandauthors/>.  
goodreads:

<http://www.goodreads.com/>.

LibraryThing:

<http://www.librarything.com/>.

PoemHunter.com:

<http://www.poemhunter.com/>.

The Poetry Foundation's

[\(http://www.poetryfoundation.org/\)](http://www.poetryfoundation.org/)

Poetry Tool:

<http://www.poetryfoundation.org/archive/poetrytool.html>.

Yahoo! directory of literary magazines:

[http://dir.yahoo.com/arts/humanities/literature/news\\_and\\_media/magazines/](http://dir.yahoo.com/arts/humanities/literature/news_and_media/magazines/).

## Appendix B:

### Potential Metadata Elements from RQ2 by Set

#### *Bibliographic:*

A library classification number such as a Dewey Decimal number or Library of Congress classification.

A library-assigned subject heading.

Information about derivative works based on the work in question, e.g., a children's version of a novel.

Information about earlier works on which a work is based, e.g., folklore on which a current story is based.

The names of libraries which hold the journal or book in which the work appeared.

The number of editions of a book that have been published.

The number of libraries which hold the journal or book in which the work appeared.

#### *Faceted:*

Factual information available in the work, e.g., the actions of historical figures.

Information about characters in a literary work, e.g., a character's religion or occupation.

Information about events in a literary work, e.g., physical actions by characters, or larger phenomena such as warfare.

Information about the era depicted in a literary work, e.g., the period or age in which actions take place, or the time of day described in the work.

Information about the geographic locations in a literary work, e.g., locations of characters or events in a fictional world.

Relationships between characters, e.g., mother-daughter.

The names of known fictional characters.

#### *Popular:*

A description of the theme of the work, e.g., Nature or Truth.

A sample of the work's content.

An indication of the author's intent for the work, e.g., humor.

An indication of the intended emotional experience of the work, e.g., tearjerker.

Comments from other readers.

The intended audience for a work, e.g., popular vs. literary.

The specific genre into which a work falls, e.g., mystery or adventure.

#### *Shared:*

An abstract or similar content description.

The age of the intended audience.

The author's nationality or ethnic background.

The broad form which a work takes, e.g., short story, poem, collection, or novel.

The broad genre into which a work falls, e.g., poetry or fiction.

The language in which the work is written.

The length of the work.

The name of the author.

The title of the work.

The year in which a work was published.

#### *Symbolic Capital:*

A list of authors who have also had books published by a particular book publisher.

A list of authors whose work appeared in a previous edition of the same anthology.

A list of authors whose work appeared in previous issues of the same literary magazine.

A list of authors whose work appeared in the same anthology.

A list of authors whose work appeared in the same issue of a literary magazine.

The name of an award given to the author.

The name of an award given to the work.

The name of the editor or editors who decided to publish the work.

The name of the publisher of the book or magazine in which the work appeared.

The title of a literary magazine in which the work has appeared.

The title of an anthology in which the work has appeared.

*Other:*

Are there other characteristics that you would use in deciding whether to read a piece of new literary work? [If Yes] Can you tell me what those characteristics are?