

The Concept of a Work in WorldCat: An Application of FRBR

Rick Bennett

Office of Research
OCLC Online Computer Library Center, Inc.
6565 Frantz Road
Dublin, Ohio 43017
bennetr@oclc.org

Brian F. Lavoie

Office of Research
OCLC Online Computer Library Center, Inc.
6565 Frantz Road
Dublin, Ohio 43017
lavoie@oclc.org

<<Please address correspondence to this author>>

Edward T. O'Neill

Office of Research
OCLC Online Computer Library Center, Inc.
6565 Frantz Road
Dublin, Ohio 43017
oneill@oclc.org

Abstract: This paper explores the concept of a work in WorldCat, the OCLC Online Union Catalog, using the hierarchy of bibliographic entities defined in the *Functional Requirements for Bibliographic Records (FRBR)* report. A methodology is described for constructing a sample of works by applying the FRBR model to randomly selected WorldCat records. This sample is used to estimate the number of works in WorldCat, and describe some of their key characteristics. Results suggest that the majority of benefits associated with applying FRBR to WorldCat could be obtained by concentrating on a relatively small number of complex works.

Keywords: Work, FRBR, WorldCat, Bibliographic Record, Descriptive cataloging

1. Introduction

The concept of a *work* is “an essential component of modern catalogs” [1]. And yet, much ambiguity surrounds its definition, particularly, as Smiraglia observes, in regard to “the degree to which change in ideational or semantic content represents a new work” [2]. *Functional Requirements for Bibliographic Records* (FRBR) [3], an initiative sponsored by the International Federation of Library Associations and Institutions (IFLA) Section on Cataloging, extends much of the previous scholarship on the nature of a work into a functional concept suitable for implementation in library catalogs. By offering a definition of a work as well as a prescription both for distinguishing *between* works and clustering together variations of a single work, FRBR represents a valuable tool for identifying, describing, and comparing works.

The FRBR model has generated a great deal of interest in the library community, with several initiatives currently underway to apply the FRBR concepts to library catalogs. In this paper, the FRBR concept of a work is applied to a sample of records taken from WorldCat (the OCLC Online Union Catalog) to: 1) estimate the number of works represented by the nearly 50 million records in WorldCat, and 2) identify the salient characteristics of these works. This paper provides a brief overview of FRBR, a description of a methodology for applying the FRBR work concept to a sample of 1,000 bibliographic records taken from WorldCat, and estimates of the number of works in WorldCat and their associated characteristics, based on analysis of the sample. Application of the concept of a work to a union catalog, in terms of its impact on the cataloging process, is briefly discussed.

2. Overview of FRBR and the concept of a work

Rapid changes in the cataloging environment, i.e., increased volume of published information and automated cataloging functions, the expectations of users of library services, and the perceived need to reduce cataloging costs have underscored the need for corresponding changes in cataloging practice. A 1990 IFLA-sponsored Seminar on Bibliographic Records, held in Stockholm, examined “the purpose and nature of bibliographic records and the range of needs that they can realistically be expected to meet and ...[considered] alternative ways of meeting those needs in a cost-effective and co-operative manner” [4]. The Seminar produced seven resolutions, one of which called for a study “to define the functional requirements for bibliographic records in relation to the variety of user needs and the variety of media.” [5] An international study group formed to address this task issued its final report in 1998: *Functional Requirements for Bibliographic Records*, or FRBR.

The definition of a work and its relationships with other bibliographic entities are essential elements of the FRBR model. Smiraglia [6] provides a detailed treatment of the concept of a work, tracing the evolution of its definition. Svenonius [7] credits the publication of Tillet’s 1987 dissertation [8] as the catalyst for later research activity exploring the nature of bibliographic relationships. Tillet [9] also provides a taxonomy of bibliographic relationships.

A number of sources have considered the potential benefits of moving FRBR from theory to practice. Noerr, et al. [10] provides an excellent discussion of this topic.

This document is an e-print version of an article published in *Library Collections, Acquisitions, and Technical Services* 27,1 (Spring 2003). The e-print file was posted on 28 March 2003 at http://www.oclc.org/research/publications/archive/2003/lavoie_frbr.pdf.

Please cite the published version.

They conclude that FRBR's primary benefits extend from its hierarchical structure, permitting the placement of bibliographic information at its appropriate level of abstraction and facilitating its inheritance at lower levels. This yields a data model that is easier to maintain, is more flexible in terms of representing cataloged materials, and offers improved searching and clustering strategies.

The architects of FRBR sought to develop a conceptual framework matching common tasks performed by users of bibliographic records to the bibliographic data necessary to fulfill them. FRBR's core insight is that a set of *entities* can be identified which are key to the successful use of bibliographic records, e.g., a *work*, a *person*, or an *event*. These entities are related to one another in a variety of ways—e.g., a work *may be created by* a person, or an event *may be the subject of* a work. Finally, each entity is characterized by a set of *attributes*. A work, for example, may be defined by a title, creation date, context, etc.; a person may have a name, title, birth and/or death date, etc. This approach emphasizes not individual data elements in the bibliographic record *per se*, but rather the entities, relationships, and attributes the bibliographic record is intended to describe. Implementation of the FRBR model in a library catalog would be expected to bring several benefits, including the ability to: 1) accommodate various user needs by supporting different views of the bibliographic database; 2) enhance retrieval through the representation of a hierarchy of bibliographic entities in the catalog (e.g., by collapsing near-duplicate items to a single entry point); and 3) increase cataloging productivity (e.g., by merging information from multiple bibliographic records so that the original or copy cataloger can select the most appropriate information for inclusion in a new record).

FRBR identifies three classes of entities relevant to users of bibliographic information: Group 1 entities include the “products of intellectual or artistic endeavor that are named or described in bibliographic records” [11]; Group 2 are “those responsible for the intellectual or artistic content, the physical production and dissemination, or the custodianship of the entities in the first group” [12]; and Group 3 entities “serve as the subjects of intellectual or artistic endeavor” [13]. *The class relevant to the present study is Group 1, which includes:*

- **Work:** a distinct intellectual or artistic creation
- **Expression:** the specific form that a work takes each time it is “realized”
- **Manifestation:** the physical embodiment of an expression of a work
- **Item:** a single exemplar of a manifestation

This four-level bibliographic structure begins with an abstract entity called a *work* at the top of the hierarchy, and runs through three levels of ever-increasing concreteness ending with the *item* entity, which refers to a single copy of a resource such as a book or CD-ROM. Each of these entities is described in greater detail in the following paragraphs.

According to FRBR, a *work* is a distinct artistic or intellectual creation by a person, group, or corporate body, which is identified by a name or title. Although the concept of a work is necessarily abstract, FRBR provides a set of guidelines for

determining the boundaries of a work in practice. Modifications involving “a significant degree of independent artistic or intellectual effort” [14] are sufficient to produce a new work. Examples of new works include paraphrases, adaptations for children, parodies, musical variations on a theme, dramatizations, adaptations from one medium to another, abstracts, digests, and summaries.

An *expression* is “the specific intellectual or artistic form that a work takes each time it is ‘realized.’” [15] The form of a work might be “alpha-numeric, musical or choreographic notation, sound, image, movement, or any combination of such forms.” [16] The key difficulty in working with the FRBR bibliographic entities lies with the concept of an expression. The stipulation that “any change in artistic or intellectual content ... no matter how minor” [17] is considered to be a new expression presents serious implementation issues. For example, determining whether or not one edition of a book represents a different expression compared to another edition can be an arduous process. The revisions or modifications, if any, may not be evident from the bibliographic record itself and, therefore, would require manual examination of the book to identify, a task which may be unrealistic or even impossible. See O’Neill [18], for a case study illustrating these problems.

A *manifestation* is “the physical embodiment of an expression of a work.” [19] Manifestations take the form of manuscripts, books, periodicals, maps, posters, sound recordings, films, video recordings, CD-ROMS, or multimedia kits—“all the physical objects that bear the same characteristics, in respect to both intellectual content and

physical form.” [20] These characteristics are those that appear at the time of manufacture; idiosyncratic attributes, such as a missing page or an autograph by the author, are not considered characteristics of a manifestation. Determining the boundaries between one manifestation and another, therefore, requires a comparison of the objects’ intellectual content *and* physical form. Examples of changes in physical form include typeface, typesetting, page layout, change from paper to microfilm, or change from cassette to cartridge. Changes in the artistic or intellectual content result in a new manifestation of a new expression of the work.

Finally, an *item* is a single exemplar of a manifestation.

In this study, the FRBR concepts of *work* and *manifestation* are used to examine the number and characteristics of works present in WorldCat.

3. Identifying work clusters in WorldCat

A random sample of 1,000 bibliographic records was selected from WorldCat. No restrictions were placed on the type of material to be included in the sample, so the distribution of the sample records across type reflects the overall distribution in WorldCat as a whole—85% books, 5% serials, 4% musical performances and scores, 3% projected mediums, 2% maps, and the remainder a variety of forms such as voice recordings, computer files, and two-dimensional non-projectable graphics.

An examination of the sample records revealed that four were associated with the Bible. Because sacred works pose unique challenges in terms of identifying their boundaries, and warrant separate study, these four records were excluded from the analysis, for a sample total of 996 records.

A WorldCat record describes the FRBR manifestation entity that, according to the structure of the FRBR model, can be traced back to the work entity from which it was derived. Therefore, the 996 sample records can also be considered a sample of works. However, since multiple manifestations can be associated with the same work, characterization of the works present in WorldCat requires first identifying any additional records in WorldCat corresponding to any of the works represented in the sample.

The process of clustering WorldCat records associated with the works in the sample was a combination of automated scans and manual review. First, WorldCat was scanned through an algorithm that utilized critical information from each sample record's main entry and title fields to identify candidate records for the cluster. For example, an author from a sample record "Smith, John Jacob" was matched to potential variations on the name, such as "Smith, John" or "Smith, J." Obvious mismatches, such as "Smith, Joseph," were excluded. In addition to author matching, records were selected based on full or partial keywords extracted from the sample record's title. Keywords were manually selected on the basis of relevance and uniqueness, and were compared to text in any title or note field. Partial keywords were particularly useful for picking up plurals, or

titles in other languages. For example, William Buchan's *Domestic Medicine* can be found in both French and Spanish using the partial keyword "domest".

The automated scan of WorldCat provided a broad capture rate for potential records associated with the work in question. The list of candidate records for each work in the sample was then reviewed manually, and these records were supplemented by *ad hoc* manual searching using OCLC's FirstSearch to investigate other variations in authors or titles not captured by the automated scan. The manual review, which confirmed that the automated scan usually captured all of the related records, therefore served primarily to discard unrelated records captured by the automated scan rather than to add new records to the list of related records.

4. Results and analysis

Creation of the work clusters as described above resulted in the extraction of an additional 7,702 records from WorldCat, for a total of 8,698 records associated with 996 sampled works. These records can be used to estimate the number of works in WorldCat and to characterize their attributes.

Prior to drawing inferences from the sample data, an adjustment had to be made to correct for any bias. Since works were indirectly selected by sampling *manifestations* from WorldCat, works with larger numbers of manifestations had a greater likelihood of being selected. This introduces bias into the sample of works, since large works (i.e., works with a large number of manifestations) would be over-represented. Since a work

This document is an e-print version of an article published in *Library Collections, Acquisitions, and Technical Services* 27,1 (Spring 2003). The e-print file was posted on 28 March 2003 at http://www.oclc.org/research/publications/archive/2003/lavoie_frbr.pdf.

Please cite the published version.

with n different manifestations has n times the probability of being selected, the observed frequency of a work of size n must be divided by n to obtain an unbiased estimate of the actual frequency. For example, if works with five manifestations were observed 22 times in the sample, this result was divided by the work size to yield a weighted frequency of 4.4. This procedure equalized *ex post* the probability of selection across works of unequal size, thereby removing inferential bias.

4.1. General statistics

As of December 2001, WorldCat contained 46,767,913 records [rounded to 47 million]. For the purposes of this study and in line with FRBR model definitions, it is assumed that each bibliographic record in WorldCat describes a *manifestation*. Based on the analysis of the sample, these 47 million manifestations can be traced back to approximately 32 million distinct *works* in WorldCat. The average work in WorldCat has approximately 1.5 manifestations, indicating that for the most part, works in WorldCat are small, single-manifestation entities. More than 25 million of the 32 million works in WorldCat (78%) consist of a single manifestation. Ninety-nine percent (99%) of all works in WorldCat have seven manifestations or less, and only about 30,000, or 1% have more than 20 manifestations.

Initial observations would suggest that the benefits of using the FRBR model to organize and improve search and retrieval functions for large works are confined to a relatively small segment of the library catalog, since works with only a single

manifestation represent trivial cases within the FRBR bibliographic entity hierarchy. If findings are interpreted in this manner, the potential scope for applying FRBR is reduced to approximately 20% of all works in WorldCat, i.e., those containing two or more manifestations. This 20% proportion can likely be narrowed even further, since FRBR yields its greatest utility for relatively large works—only 1% of all works in WorldCat contain eight or more manifestations.

In no way is this interpretation—paring down the potential benefits to 1% of all works—meant to understate the potential of FRBR. Consider the following:

- One percent of 32 million works in WorldCat projected through the sample, is 320,000 works, which, in absolute terms, is still a significant number. As a point of comparison, consider that the average Borders bookstore contains 150,000 books [21]. These books correspond to the FRBR concept of an item. Assuming multiple copies are kept in stock, these items can be traced back to a proportionately smaller number of manifestations. These manifestations correspond, in turn, to a smaller number of expressions, and ultimately, to an even smaller number of works. Therefore, the number of works represented in Borders will be some small fraction of 150,000. Given this, applying FRBR to 1% of the works in WorldCat, or 320,000 works, would account for many times the number of works found in a large bookstore such as Borders.

- Persuasive evidence can be marshaled to support the hypothesis that the largest works represent the most important segment of the catalog, as measured by library holdings data. For example, the five most widely held works in the WorldCat sample, as measured by total holdings, were also the five largest works. The most widely held work in the sample that had just one manifestation exhibited total holdings of 710. In contrast, the largest work in the sample (1,251 manifestations) had total holdings of 27,434. These data suggest that applying FRBR to a small segment of the library catalog, i.e., the largest works, would yield a disproportionately high degree of benefit for the most libraries.

4.2. *Types of work*

As discussed in Section 2's overview of the FRBR model, a work can embody multiple expressions. For these works it is useful to examine the nature of the variation that distinguishes one expression from another. Such an analysis offers insight into the complexity of the works identified in the sample, as well as the dynamic evolution of a work over time, relative to its original expression.

To conduct this analysis, the authors defined three classes of works:

- An **Elemental Work** is a work with a single expression and a single manifestation, such as a government report that was published exclusively as a pamphlet.

- A **Simple Work** is a work with a single expression but multiple manifestations, such as a doctoral thesis available in both paper and microfilm.
- A **Complex Work** is a work with multiple expressions, or realizations, of its intellectual or artistic content, such as multiple editions of a textbook.

Figure 1 illustrates the breakdown of works in WorldCat by degree of complexity. Elemental works account for the largest proportion of works in WorldCat, followed by Simple works, and then Complex works at only 6% of all works in WorldCat.

<<FIGURE 1>>

Since they embody multiple expressions, complex works will tend to be larger than the average work in WorldCat. As the discussion above suggests, larger, complex works stand to benefit the most from organization under the FRBR entity-relationship model.

The challenge comes in identifying complex works. This requires a means to identify the existence of multiple expressions subsumed under a single work or, put another way, the *forms of variation* that distinguish one expression from another. According to the FRBR model, “*any* [emphasis added] change in intellectual or artistic content constitutes a change in expression” [22]. A strict interpretation of this definition

implies that even relatively minor changes, such as an updated bibliography, are sufficient to create a new expression of the work and thus define the work as complex.

In some cases, multiple expressions within a work are straightforward to identify. For example, multiple translations of a particular work can be easily identified from information present in the bibliographic record. Other forms of expressions constitute subtler variations in content that may not be discernable from data in the records. (See O'Neill [23] for a case study in identifying expressions.) In these cases, manual inspection of the physical items of the expressions are required to determine if a work is complex, i.e. has multiple expressions and manifestations.

The 996 sample works from WorldCat were examined to gain insight into the scope for identifying and categorizing various types of expressions embodied in complex works, using only information available in the bibliographic records associated with a work. The records for each work in the sample were manually reviewed in order to identify patterns or commonalities useful for characterizing distinct categories of complex works. From this analysis, six categories emerged:

- **Augmented Works:** intellectual or artistic content is supplemented by additional material: e.g., illustrations, prefaces, etc.

Example: Smart, Christopher: *Jubilate Agno*

Expressions:

undated manuscript (the author died in 1771)

1939 version, edited by Stead, William Force

1954 version, with introduction and notes by Bond, W. H.

1965 version, illustrated by Baskin, Lisa Unger

1980 version, with afterword by Heckscher, Philip Hofer

1996 selection from the work, no supplemental materials

- **Revised Works:** intellectual or artistic content is revised; typically, current version supercedes previous versions

Example: Ollard, E.A. and E.B. Smith: *Handbook of Industrial Electroplating*

Expressions:

originally published in 1947

1954 edition

1964 edition

- **Collected/Selected Works:** any combination of multiple works by a single author

Example: Sheridan, Richard Brinsley

Expressions:

The Plays of Richard Brinsley Sheridan

A Volume of Plays: As Performed at the Theatre, Smoke-

Alley, Dublin

Complete Plays

Plays & Poems

The Dramatic Works of Richard Brinsley Sheridan

Sheridan's Plays Now Printed as He Wrote Them and His

Mother's Unpublished Comedy, A Journey to Bath

Six Plays

The Humorous Plays of Richard Brinsley Sheridan

(Œuvres Dramatiques du Très Honorable Richard

Brinsley Sheridan)

- **Multiple Translations:** intellectual or artistic content is unchanged, but is represented using multiple intellectual conventions and instruments (e.g., languages).

Example: Novak, Vaclav: A Short History of Czechoslovakia

Expressions:

A Short History of Czechoslovakia

Compendio Historico de Checoslovaquia

Breve Storia della Cecoslovacchia

Krotki Zarys Historii CSRS

- **Multiple Forms of Expression:** intellectual or artistic content is expressed using multiple forms of expression (e.g., text, images, sound, etc.)

Example: Halpern, Shari: My River

Expressions:

presented as printed text

presented as a sound recording

- **Multiple Translations, Multiple Forms of Expression:** both multiple translations, and multiple forms of expression are embodied in the work.

Collected/selected works (third category above) are a special case, in that they are difficult to fit into a strict interpretation of the FRBR model. According to FRBR, works “may represent an aggregate of individual works brought together by an editor or compiler in the form of an anthology, a set of individual monographs brought together by a publisher to form a series, or a collection of private papers organized by an archive as a single fond.” [24] For the purposes of this study, this FRBR definition was broadened to include any aggregation of works by a single author. Variations in the set of works constituting the aggregation are then considered a different expression of the same work. For example, “the collected works of Shakespeare” would be considered a work; a collection of Shakespeare’s comedies and a collection of Shakespeare’s tragedies would be considered two distinct expressions of this work.

It should be noted that none of these six categories defined above are mutually exclusive. In assigning a work to a category, precedence was given to the augmented, revised, or collected/selected categories. Only if a work fell outside the bounds of these categories were the other three categories considered. Also, categorization was based

strictly on information available in the bibliographic records. In general, augmented works were identified by information in the 700 field (added entries); revised works by the 250 field (edition statement); and collected works by the 245 field (title). Translations were determined on the basis of information in the 008 field, while forms of expression were derived from the “Type” and “Bibliographic Level” positions of the record leader.

As noted above, complex works make up approximately 6% of all works in WorldCat, or a little less than 2 million works. Figure 2 illustrates the breakdown of complex works by type.

<<FIGURE 2>>

Based on analysis of the sample, it is estimated that more than half of the approximately two million complex works in WorldCat are revised works; a quarter embody expressions distinguished solely on the basis of language, and the remaining categories account for relatively small segments of complex works (less than 10 % each).

Although complex works account for only a relatively small percentage of the works in WorldCat, this result belies the true significance of these works. For example, revised, augmented, and collected/selected works together account for only 4% of the works in WorldCat. Yet, these works represent more than 12% of the manifestations (records). This suggests that works falling into these categories will tend to be larger, in terms of number of manifestations, than the average work in WorldCat. Indeed,

augmented works contain, on average, approximately 15 manifestations, while revised works and collected/selected works each contain, on average, about four manifestations.

In addition to accounting for a disproportionately large portion of the records in WorldCat, complex works also represent a relatively high percentage of the most widely held works (based on total holdings). For example, the top twenty most widely held works in the sample are all complex works; all but three of the top fifty are complex. This suggests that the application of FRBR to library catalogs might usefully begin by concentrating on complex works. Because these complex works constitute a small proportion of all works, the scope of the task is manageable while yielding the greatest benefits.

4.3. Characteristics of works in WorldCat and the impact on cataloging

In addition to improving search and retrieval functionality for users of bibliographic records, FRBR also creates the potential for realizing “economies of scale” in cataloging. This occurs by propagating characteristics applicable at the level of a work among all manifestations of that work. Put another way, these characteristics apply to the work as a whole, and, therefore, are “inherited” by all manifestations associated with a particular work.

To explore this aspect of FRBR, the sample works from WorldCat were examined in regard to *subject*. Assigning subject headings and classification numbers is a time-

consuming and expensive process; therefore, this characteristic is particularly important in regard to the notion of inherited bibliographic data. For the purposes of this study, collection of information pertaining to the subject of a work was confined to what was available in the bibliographic record, rather than through physical inspection of a manifestation of the work. Information was parsed from the 050 and 090 fields (Library of Congress classification numbers), the 082 and 092 fields (Dewey classification numbers), and the 600 – 651 fields with second indicator equal to zero (Library of Congress Subject Heading).

Table 1 characterizes the use of classification numbers in regard to works in WorldCat. Figure 3 following the table shows the proportion of works with classification numbers, by size.

<<TABLE 1>>

<<FIGURE 3>>

As the table shows, close to 17 million of the 32 million works in WorldCat (53%) contain at least one record with a Library of Congress classification number. In comparison, about 8.5 million works (27%) contain a Dewey Decimal classification number, and just more than 5 million works (17%) contain both Library of Congress and Dewey classification numbers. Surprisingly, more than a third of all works in WorldCat lack a record with either an LC or a Dewey number.

A direct correlation is discernable between the number of manifestations associated with a work and the likelihood that at least one of these manifestations will contain a classification number in its bibliographic record. According to the sample data, if there is a single manifestation associated with a work, the chance of it containing a Library of Congress classification number is about 50%. When two manifestations are associated with a work, the probability of obtaining at least one LC number increases to about 67%. The probability increases still further to 85% if there are three manifestations. For works with 4, 5, 6, 7, or 8 manifestations, the chances of obtaining an LC classification number are between 90 and 95%. For works embodying nine or more manifestations, obtaining an LC classification number from at least one record is virtually certain.

Dewey classification numbers are less common than LC numbers. For works of one manifestation, there is only a 22% chance of containing a Dewey number; this probability increases to 37% and 53% for works with two or three manifestations respectively. For works with 15 or more manifestations, it is virtually certain that at least one record will contain a Dewey number.

Clustering manifestations into works permits the inheritance of certain types of bibliographic information across all bibliographic records associated with a work. Work-level information that appears in only a few, or even one, records can be extended to all records in the cluster. For example, the analysis of sample data would suggest that about

48% of the records in WorldCat contain an LC classification number. This proportion, however, increases to 53% when *works* are considered. These 17 million works account for about 30 million WorldCat records, of which 23 million contained an LC classification number in the 050 or 082 fields. Given that records associated with the same work should share the same classification number, all 30 million records associated with the 17 million works therefore possess an LC number, either explicitly by the inclusion of the 050 or 082 fields, or implicitly, through an association with another record containing this data and matched to the same work. The result is an increase by one third in the number of WorldCat records with an LC number.

This effect is even more pronounced with Dewey classification numbers. In this case, there are 8.5 million works in WorldCat containing at least one record with a Dewey number. These works embody 17.5 million manifestations. Of these, 11 million have a WorldCat record that explicitly contains a Dewey number. The rest, however, can inherit this information, resulting in an increase of over 50% in the number of WorldCat records containing a Dewey number.

The propensity for work clusters to contain one or more Library of Congress subject headings was also examined. Analysis of the sample indicates that a little more than three-quarters of the works in WorldCat contain at least one LC subject heading among the bibliographic records matched to each work. The average work in WorldCat contains approximately 2.3 subject headings, distributed among one or more of the bibliographic records constituting a particular work cluster. This number drops

significantly, however, when only unique subject headings are considered: in this case, the average work in WorldCat contains 1.8 subject headings.

The analysis of subject headings in the sample records highlights a key challenge in leveraging work-level information across all manifestations—consistency in providing values for this information. For example, there is no reason why the number of unique subject headings associated with a work should increase as the number of manifestations increases. This supposition, however, is belied by the sample data, where a loose, positive correlation between work size and the number of unique subject headings is discernable. This correlation suggests that different catalogers are assigning different subject headings to the same work when creating the bibliographic record for their particular manifestation. Since subject headings are primarily used as access points in the catalog, an argument can be made that more subject headings are beneficial, since the probability that the work will be “discovered” by users is increased. It is more likely, however, that a smaller set of widely agreed-upon subject headings would be preferred.

The ability to leverage the information embodied in a set of work records will be greatly influenced by the design of the cataloging system. Key issues include the selection of which work-level information is made available to the cataloger, the design of the cataloging interface, and the ability to create cataloger profiles so that the appropriate information is displayed at the beginning of the cataloging session. These issues will be of increasing importance as more non-AACR2 records are added to WorldCat.

5. Conclusion

FRBR is a significant contribution to cataloging on both theoretical and practical grounds. Theoretically, FRBR proposes a comprehensive bibliographic framework that defines the key entities of interest to users of bibliographic records, enumerates their salient attributes, and articulates the various relationships existing between these entities. Practically, the FRBR entities and relationships lend themselves to implementation in library catalogs as functional concepts, designed to improve the utility of bibliographic records as tools for both reference and cataloging. A major component of FRBR's contribution in both of these areas is its examination of works and their role in the library catalog: "The [FRBR] report," observes Smiraglia, "represented a major milestone in the history of the treatment of works in catalogs by defining them in concrete terms and by providing an entity-relationship schema for their deliberate incorporation into catalogs ... [T]he report reversed the functional emphasis of item over work that had been characteristic of catalog construction heretofore." [25]

The application of FRBR to WorldCat, the world's largest union catalog, demonstrates several potential benefits in library catalogs. First, the sample data suggests that the task of applying FRBR may not be as burdensome as *a priori* estimates might suggest: FRBR can be applied non-trivially to only a small percentage of works in WorldCat. At a maximum, 20% of the works would be candidates (i.e., works with two or more manifestations); in practice, however, the percentage is likely to be much lower.

Analysis suggests that concentrating on relatively large works, in particular those works whose content has been augmented, revised, or consists of collections of other works (a relatively small portion of the catalog) might be sufficient to capture the lion's share of benefits potentially available from implementing FRBR.

The difficulty in applying FRBR to library catalogs would be eased by the availability of algorithms to perform at least part of this task through machine processing of bibliographic records. A study by Hickey, et al. [26] offers some promising results in this regard.

Analysis of the sample also suggests that FRBR may serve as a means of leveraging information in particular bibliographic records across other records in the catalog, reducing the cost and increasing the quality of both original and copy cataloging. These benefits are obtained even if a local system does not fully incorporate a FRBR structure. The structure of the FRBR model implies the existence of certain information that applies at the highest level of the bibliographic hierarchy – the work – and therefore also applies to, or is inherited by, bibliographic entities comprising the lower levels of the hierarchy. This includes manifestations, the entity represented by a WorldCat record. Given this, work-level information – such as classification numbers or subject headings – can be propagated amongst all manifestations associated with a particular work, even though the information may have been explicitly recorded in only one record. In this way, the aggregation of records into clusters associated with the FRBR concept of a work permits the realization of further economies of scale in cooperative cataloging.

The FRBR model, with its definition of concepts and relationships associated with bibliographic entities, promises to improve the functionality of search and retrieval tools for catalog users, as well as introduce greater efficiencies in cataloging practice. The analysis of the sample of works in WorldCat reported in this study suggests that these benefits do in fact exist, and could be obtained in large part through “FRBRizing” a relatively small portion of the catalog. More research needs to be done, however, to examine the issues and challenges associated with implementation of the FRBR model in library catalogs. A key area for further work is the need to transform the conceptual definitions of the FRBR entities into clear, implementation guidelines. Although it is unlikely that identification of FRBR entities can be unambiguous, more precise delineations between the four FRBR levels would facilitate their application. More “FRBRization” case studies would also assist in understanding the implementation process. With working definitions of the entities in hand, improved algorithms for identifying these entities from bibliographic records will be possible, diminishing the burden of applying FRBR retrospectively.

References

- [1] Smiraglia, R.P. (2001) The “nature” of a work: implications for the organization of knowledge. Lanham: Scarecrow, p. 15
- [2] Ibid, p. 52
- [3] IFLA Study Group on the Functional Requirements for Bibliographic Records (1998) Functional requirements for bibliographic records: final report. Munchen: KG Saur.
- [4] Bourne, R., editor (1992) Seminar on bibliographic records: proceedings of the seminar held in Stockholm, 15-16 August 1990, and sponsored by the IFLA UBCIM Programme and the IFLA Division of Bibliographic Control. Munchen: KG Saur, p.2
- [5] Ibid, p. 145
- [6] Smiraglia (2001)
- [7] Svenonius, E. (2000) The intellectual foundation of information organization. Cambridge: MIT Press, p. 100
- [8] Tillet, B.B. (1987) Bibliographic relationships: toward a conceptual structure of bibliographic information used in cataloging. Ph.D. dissertation: University of California, Los Angeles
- [9] Tillet, B.B. (1991) A taxonomy of bibliographic relationships. *Library resources and technical services* 35(2):150-158
- [10] Noerr, P., Goossens, P., Matei, D., Otten, P., Peruginelli, S., and Witt, M. (1998) User benefits from a new bibliographic model: follow-up of the IFLA functional requirements study. 64th IFLA general conference. Available online at: <http://www.ifla.org/IV/ifla64/084-126e.htm>

- [11] FRBR, p. 12
- [12] Ibid
- [13] Ibid
- [14] Ibid, p. 17
- [15] Ibid, p. 18
- [16] Ibid
- [17] Ibid, p. 19
- [18] O'Neill, E.T. (2002) FRBR: application of the entity-relationship model to *Humphry Clinker*. (submitted for publication)
- [19] FRBR, p. 20
- [20] Ibid
- [21] Varian, H., and Shapiro, C. (1998) Information rules: a strategic guide to the network economy. Cambridge: Harvard Business School Press.
- [22] FRBR, p. 19
- [23] O'Neill (2002)
- [24] FRBR, p. 28
- [25] Smiraglia, p. 48
- [26] Hickey, T.B., O'Neill, E.T., and Toves, J. (2002) Experiments with the IFLA functional requirements for bibliographic records (FRBR). D-Lib magazine 8(9). Available online at: <http://www.dlib.org/dlib/september02/hickey/09hickey.html>

Figure 1: Works in WorldCat, By Type

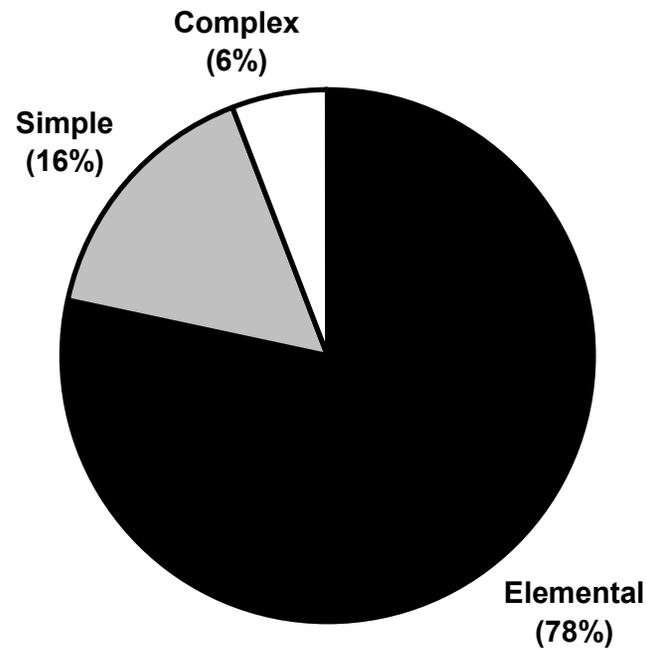


Figure 2: Types of Complex Work

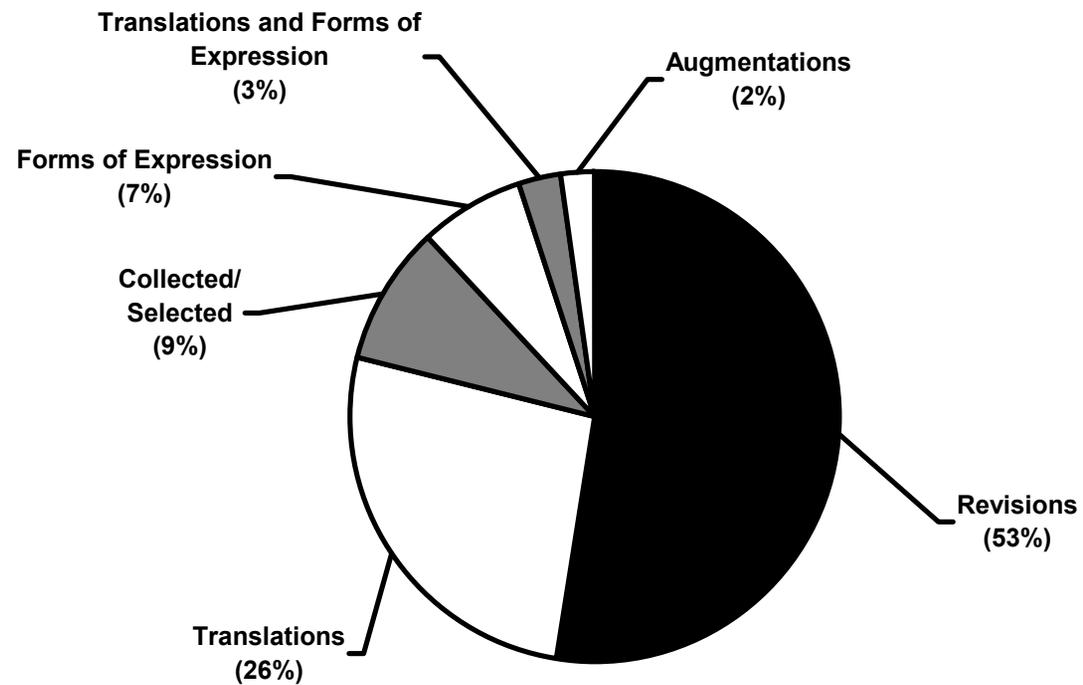


Table 1: Use of Classification Numbers in WorldCat Works

	Number	Proportion of Works
Works with LC classification number:	16,985,138	53%
Works with Dewey Decimal number:	8,527,530	27%
Works with both LC and DDC numbers:	5,389,322	17%
Works with no classification number:	11,832,851	37%

Figure 3: Proportion of Works with Classification Number, By Size

