FRBR: Functional Requirements for Bibliographic Records

Application of the Entity-Relationship Model to

Humphry Clinker

Edward T. O'Neill

Research Scientist
Office of Research
OCLC Online Computer Library Center, Inc.
6565 Frantz Road
Dublin, Ohio 43017-3395
614-764-6074
oneill@oclc.org

FRBR: Application of the Entity-Relationship Model to *Humphry Clinker*

The report from the IFLA Study Group on the Functional Requirements for Bibliographic Records (FRBR) recommended a new approach to cataloging based on an entity-relationship model. This study examined a single work, The Expedition of Humphry Clinker, *to determine benefits and drawbacks associated with creating such an entity-relationship model.* Humphry Clinker *was selected for several reasons—it has been previously studied, it is widely held, and it is a work of mid-level complexity. In addition to analyzing the bibliographic records, many books were examined to ensure the accuracy of the resulting FRBR model. While it was possible to identify works and manifestations, identifying expressions was problematic. Reliable identification of expressions frequently necessitated the examination of the books themselves. Enhanced manifestation records where the roles of editors, illustrators, translators, and other contributors are explicitly identified may be a viable alternative to expressions.   For* Humphry Clinker*, the enhanced record approach avoids the problem of identifying expressions while providing similar functionality. With the enhanced manifestation record, the three remaining entity-relationship structures—works, manifestations, and items—the FRBR model provides a powerful means to improve bibliographic organization and navigation.*

Introduction

The report from the IFLA Study Group on the Functional Requirements for Bibliographic

Records (FRBR) includes a recommendation for a fundamentally new approach to cataloging

(IFLA 1998, 13). It proposes an entity-relationship model, with four primary entities—*work,*

expression, manifestation, and *item*—representing the products of intellectual or artistic

endeavor. This shift in cataloging focus requires not simply describing the item in hand but also

describing how the item relates to other members of its bibliographic family. La Boeuf

recognizes that FRBR "is likely to induce profound changes in cataloguers' landscape" (2001,

15).

The FRBR model defines three distinct groups of entities (IFLA 1998, 12):

1. The products of intellectual or artistic endeavor (a publication)
2. Those responsible for the intellectual or artistic content (person or corporate body)
3. Those that serve as the subjects on intellectual or artistic endeavor (concept, object, event, and place)

This study focuses on the Group 1 entities. While these entities represent only one aspect of

FRBR, they are the foundation of the model.

The FRBR model proposes four entities in Group 1: *Works, Expressions, Manifestations,* and

Items. Figure 1, adapted from the corresponding figure in the IFLA report, illustrates the

relationships between these four entities (IFLA 1998, 13). Current cataloging practice focuses

on a single bibliographic unit: the physical manifestation. The FRBR model, by contrast,

proposes this four-level hierarchical bibliographic structure. Tillett (2001, 31) points out that, with the entity-relationship cataloging model, "The opportunity exists to move beyond the current 'record' structure and beyond relational and even the current object-oriented databases." However, the FRBR model requires that the bibliographic items be analyzed in greater detail to relate them to the other members of the work.

The four Group 1 entities represent two different aspects of user interest, the intellectual endeavor and the physical manifestation. The IFLA report (IFLA 1998, 12) defined each of the entities: "W*ork* (a distinct intellectual or artistic creation) and *expression* (the intellectual or artistic realization of a *work*) reflect intellectual or artistic content. *Manifestation* (the physical embodiment of an *expression* of a *work*) and *item* (a single exemplar of a *manifestation*) reflect physical form."

None of the four FRBR entities are new—most have been discussed in the literature for years. More than 40 years ago, Verona (1959, 79) defined three objectives of the catalog as:

- the rapid location of a particular book [manifestation].

- the provision of information concerning all editions, translations, etc. [expressions] of a given work as far as they exist in the library.

- the provision of information concerning all works by a given author as far as they exist in the library.

Four years later, Lubetzky (1963) and Verona (1963) discussed these objectives in detail, generally agreeing that using the *manifestation* as the basic entity best served the first objective, but using the *work* as the basic entity best serves the second objective. Since the card catalog

could not support a hierarchical model, the selection of the basic entity for cataloging was an *either/or* decision. Most cataloging codes, including AACR, chose the manifestation as the basic bibliographic unit.

Since the Lubetzky and Verona discussion, technology has changed dramatically, with the online catalog replacing the card catalog. The online catalog does not have the same limitations and, thus, it is no longer an either/or choice of bibliographic unit. Online catalogs *can* support hierarchical models, thereby removing the technical barriers to implementation of an entity-relationship model such as that proposed in the FRBR model.

The IFLA report stresses that its suggested entity-relationship model is conceptual and "does not presume to be the last word on the issues it addresses" (IFLA 1998, 5). As such the discussion herein of the basic entities, while based on the FRBR model, also is heavily influenced by the other studies. Smiraglia (2001) provides a detailed review of this literature, and compares and contrasts the terminology and definitions.

A ***work*** is a product of the intellectual or artistic activity by a person, a group, or a corporate body that is identified by a normalized title and/or name. The FRBR report stresses that a *work* is an *abstract* entity, and recognizes that "the line of demarcation which lies between one *work* and another" (IFLA 1998, 16) is not unambiguous. Modifications involving a significant degree of independent intellectual effort such as paraphrases, rewritings, adaptations, parodies, abstracts, digests, and summaries are considered to be different *works*.

In the literature, the term *work* is frequently used interchangeably with *title*. The *work* has received limited recognition in cataloging codes, and the uniform title is commonly used to identify manifestations of a *work*. It is often argued that the hypothetical 'typical user' thinks in terms of titles, requesting, for example, *The Expedition of Humphry Clinker* rather than a particular edition of that *work*. Although concept of *work* is old, finding an acceptable definition has proven elusive. Svenonius (2000, 35) argues, "critical as it is in organizing information, the concept of *work* has never been satisfactorily defined."

An **expression** is the "realization of a work in the form of alphanumeric, musical, or choreographic notation, sound, image, object, movement, or any combinations of such forms" (IFLA 1998, 18). Like works, *expressions* are abstract entities: There is no physical referent for an *expression*. The boundaries of an *expression* are defined to exclude aspects of physical form such as typeface or page layout. The terms *text* and *edition* are commonly used to describe an *expression*, although they are often used in ways that differ from the FRBR definition.

Revisions, updates, abridgements, enlargements, and translations of an *expression* are considered new and different *expressions*. Conceptually, each unique *expression* of a work represents an intellectual or artistic activity intended to update, enhance, or otherwise modify the context of a work. All manifestations of an *expression* contain the identical content. However, the overall appearance and usability of these manifestations may differ significantly due to differences in the materials, design, and manufacturing process used to produce them. A microform reproduction certainly will have a different look and feel from the hand-printed leather-bound volume from which it was derived, even though their contents are identical.

A ***manifestation*** is the physical embodiment of an expression and "encompasses a wide range of

materials including manuscripts, books, periodicals, maps, posters, sound recordings, films,

video recording, CD-ROMs, and multimedia kits." As an entity, a *manifestation* represents all

the physical objects that bear the same intellectual and physical characteristics" (IFLA 1998, 20).

Changes in typeface, size of font, page layout, or change of publisher will result in a new

manifestation. New printings will not result in a new *manifestation* unless other changes are

made. A *manifestation* may have different bindings (hardcover versus paperback), types of

paper (regular or acid-free), or other variations (thumb-indexed) that do not significantly affect

the printed image. The *manifestation* is roughly the equivalent to the bibliographic item that

currently serves as the basis for most cataloging codes.

An ***item*** is single example of one, single manifestation. Changes that occur after the

manufacturing process (defacement, rebinding) are considered changes to the *item* and do not

result in a new manifestation. The *item* is a single logical unit but, not necessarily, a single

physical unit. Books published in multiple volumes, for example, are a single bibliographic *item*.

The most important aspect of the FRBR model is the relationships between the entities in a

group. "A *work* is realized through an *expression*...[or, in reverse] an *expression* is a realization

of a *work*. This relationship serves as the basis for "identifying a *work* represented by an

individual *expression* and for ensuring that all *expressions* of a *work* are linked to the *work*"

(IFLA 1998, 58-59). Similarly, an *expression* is "embodied in a *manifestation*, or conversely that

a *manifestation* is the embodiment of an *expression*." These logical connections help to identify

"the *expression* of a *work* embodied in an individual *manifestation* and for ensuring that all *manifestations* of the same *expression* are linked back to that *expression*" (59). The relationship continues by connecting *manifestation* with *item*, which is a single example of a *manifestation*.

Humphry Clinker

The goal of this study was to go beyond organizing bibliographic records to organizing the bibliographic objects represented by bibliographic records. This effort focused on:

- examining the benefits and drawbacks associated with creating an entity-relationship model for a work,

- better understanding the relationship between bibliographic records and the bibliographic objects they represent,

- determining if information available in bibliographic records is sufficient to reliably identify the FRBR entities,

- developing a data set that can be used to compare and evaluate FRBRization algorithms.

Building an FRBR entity-relationship model for a nontrivial work and studying the work in detail appeared to be the best way to meet these objectives. The work selected was *The Expedition of Humphry Clinker* by Tobias Smollett. *Humphry Clinker,* originally published in 1771, is generally considered to be Smollett's finest novel and one of the better works of eighteenth-century English fiction. The World's Classics edition of *Humphry Clinker* (Oxford University Press, 1984) provides of a brief description of the novel:

William Thackeray referred to Smollett's last novel, The Expedition of Humphry

Clinker, as 'the most laughable story that has ever been written since the goodly

art of novel-writing began'. First published in 1771, and often regarded as

Smollett's finest book, it relates, in an ingenious series of overlapping letters, the

adventures of Mr. Matthew Bramble's family party as they travel through

England and Scotland, visiting places such as Bath, London, Edinburgh, and the

Highlands. The group includes a gouty country squire, a husband-hunter, an

Oxford student, and an illiterate but racy lady's maid. They recount their

travelling adventures. They gossip. They tell stories of Humphry Clinker, a

servant picked up *en route*, and they record their individual reactions to the tour.

All is engrossing and entertaining and, at the same time, provides through the

satire and wit a vivid and detailed picture of the contemporary social and political

scene.

The novel takes the form of overlapping letters. A typical letter is shown in Figure 2.

This letter, from the semi-literate servant Tabitha Bramble, includes numerous

misspellings and other grammatical errors.

Humphry Clinker was selected for this study for several reasons:

- It has been previously studied. It was first described as a work at the Conference on the

  Conceptual Foundations of Descriptive Cataloging held at UCLA in 1987 (O'Neill and

  Vizine-Goetz, 1989). They reported that in OCLC's WorldCat there were "110

[bibliographic records for *Humphry Clinker*] records representing 53 different publishers over a 200-year time period."

- It is work of mid-level complexity—neither the most nor least important work, and neither typical nor atypical. Many other works, particularly literary works, exhibit similar attributes.

- It is widely held. There are 179 records in OCLC's WorldCat representing more than 5,000 holdings.

It was assumed that if the FRBR entity-relationship model can be successfully applied to *Humphry Clinker*, it can be successfully applied to a broad class of similar works. Conversely, if the FRBR entity-relationship model cannot adequately represent *Humphry Clinker*, there will be many other works for which the FRBR model will also be inadequate.

In December 2001, OCLC's WorldCat was searched for all possible bibliographic records for *Humphry Clinker*. Each WorldCat record was checked to see if it was attributed to an author with a name similar to "Tobias Smollett" or if it had a title similar to "The Expedition of Humphry Clinker." This initial search resulted in very high recall but low precision. Using the FRBR definition of a *work*, the results were extensively reviewed to remove records that were not part of the *Humphry Clinker* work. This resulted in 179 records being identified, including 14 records for microforms and eight records for translations. This set of 179 bibliographic records and supporting data are available for review on the project Web site (OCLC, 2002). Identifying the bibliographic records associated with *Humphry Clinker* did not pose a significant

problem. Hickey, O'Neill, and Toves (2002) found that bibliographic records contain sufficient

information to reliably identify works.

The Evolution of Humphry Clinker

Prior to FRBRizing *Humphry Clinker*, the work was studied to achieve an understanding of its

evolution. For this purpose, it would have been ideal to collect all manifestations to permit

detailed examination and side-by-side comparisons. However, this was impractical since many

of the manifestations were in rare book collections or in poor physical condition. They were

scattered over a large number of libraries with no single library holding a significant proportion

of the different manifestations.   The various manifestations had to be examined separately and

enough information captured to permit later comparisons.

To capture as much information as possible about the book examined, a digital camera was used

to photograph key pages. This proved to be very effective: It was more convenient, less

expensive, and easier on the books than using a copier that could have damaged many of the

older books. Key pages that were photographed included the title page, verso, the first page of

the text, a particular pre-selected letter, the last page, the first page of any supplemental matter,

illustrations, and other pages that could help differentiate between similar manifestations. In all,

38 books were examined and almost 600 digital photographs were taken.

After a review of the content of the bibliographic records, the examination of the books, and the

review of the digital images, it became clear that, except for the translations, the original text of

Humphry Clinker had not been significantly changed. Changes to the original text involved

correcting minor errors, repositioning the date on letters, moving chapter headings to the top of

the page, and replacing the "ſ" (the long s). *Humphry Clinker* was originally published with the

long s as in "The pills are good for nothing—I might as well ſwallow ſnowballs." The long s was

not observed in any editions published since 1800. Except for replacing the long s, most readers

would probably not notice these changes.   Applying a strict definition of *expression*, any of

these changes may be sufficient to create a new *expression*.   However, the use of the long s

could be considered as simply a typeface and, since the other errors were created during the

typesetting phase of the manufacturing process, it can be argued that they would produce a new

manifestation rather than a new *expression*.

Unlike these minor changes, the other revisions were intentional and, therefore, should be

considered different *expressions*. Most of the intentional changes occurred by *supplementing* the

original text with additional material. Clearly, some of these additions are more significant than

others. However, the addition of any supplemental material is sufficient to create a different

expression. The following additions were observed in the sample:

- Acknowledgment

- Bibliography

- Biographical note

- Adding chapter titles

- Chronological table

- Dedication

- Glossary

- Illustrations

- Introduction and/or forward

- List of illustrations

- Map(s)

- Notes

- Publisher 's note

- Table of contents

- Textual notes

- Reproduction of original title page

- Reviews

The significance of supplemental material varied considerably. Some supplemental material is relatively minor in importance, such as the dedication "To Mary, with love" (University of Georgia 1990). Other than Mary, few readers are apt to seek out this particular edition solely because of its dedication. In other cases, such as 22 pages of notes (Oxford University Press 1998), the supplemental material provides extensive assistance to the reader and some readers will seek this edition specifically for the notes. Some supplemental material, like a chronological table, could assist some readers. However, it is unlikely that many readers would seek out a particular edition because of a chronological table. Features of these types are rarely, if ever, reflected in the bibliographic records. Yet, under the strict interpretation of FRBR, the addition or change to any of this supplemental material is sufficient to create a new expression.

Introductions, forewords, notes, and other similar supplements were the most significant and were generally attributed to an editor. At least 23 different editors have contributed to *Humphry*

Clinker, only 14 of which were used as added entries in any of the bibliographic records. The other editors were identified either by looking at other fields in the bibliographic record, notably the statement of responsibility, or by physically examining the books. Even the editors who were identified in some records were not necessarily identified consistently. An editor may have been explicitly identified with an added entry in one bibliographic record but not in a different record for a book for which the editor played the identical role.

Many of the *Humphry Clinker* illustrators are respected artists, and their contributions certainly are important to some readers. As a group, these illustrators are well recognized—at least seven of the nine have established entries in the NACO name authority file. Identifying the illustrators was particularly problematic. Sixty-seven English-language bibliographic records were identified as illustrated in the physical description (300) field. While the physical description was found to be reliable, no dependable way was found to identify the particular illustrator. Less than a third of records for the illustrated editions identified the illustrator. Unless the illustrator is explicitly listed on the title page, it is unlikely that an added entry was created. As with editors, the practice of creating added entries for illustrators was inconsistent, even when the illustrator was explicitly listed on the title page.

Bibliographies are another common significant supplement and were frequently noted in a bibliography note. However, bibliographic records rarely contained sufficient information to determine if the bibliographies in different manifestations were the same, and the bibliographer was rarely identified. Three Oxford University Press editions illustrate the problem of identifying changes in bibliographies. The bibliographies from the equivalent sections of the

three editions are shown in Figure 3. Between the 1972 and the 1984 editions of *Humphrey Clinker*, the Thomas Nelson and Sons, the Dolphin Books, and the reprint of the Everyman Edition editions were dropped from the bibliography, and eight other editions were added. Details on the editors also were added. Between the 1984 and the 1998 editions, the bibliography was updated to include four new editions. The problem in identifying the differences is compounded by the fact that the last two editions had identical pagination (xxiv, 375). For readers interested in the bibliography, these updates are important but are not reflected in the bibliographic record. Even a side-by-side comparison of the 1984 and the 1998 editions initially failed to recognize that these were different expressions.

These apparent inconsistencies in the bibliographic records are a serious impediment to identifying expressions. There are a variety of reasons for these observed inconsistencies. The books were published and cataloged over several centuries under various cataloging rules and much of the cataloging occur prior to MARC, AACR2, or the common use of shared cataloging. Some aspects of AACR2 (1988) seem to contribute to the inconsistencies by emphasizing relative, rather than absolute, significance in determining when to create entries for editors, illustrators, and other contributors.

Rule 21.30A1 limits the number of contributors to three—a single entry is specified if there are four or more contributors. This "rule of three" can result in an entry for an editor being made in one case but not in another even when the editor's contribution, e.g., a foreword to both is identical.   In such cases, it is implied that the other contributors reduce the relative significance of the foreword. For *Humphry Clinker*, the rule of three's impact was significant. All of the

records had Tobias Smollett as the main entry leaving no more than two other entries available for contributors of the supplemental material.

Rule 21.30K2 provides guidance on when to make an added entry for illustrators. While there are three conditions specified in this rule, the only condition applicable to *Humphry Clinker* is that an added entry should be made if "the illustrations are considered to be an important feature of the work." In the case of *Humphry Clinker*, this rule is difficult to apply consistently. Since the majority are not illustrated, it is difficult to argue that the illustrations are an essential feature. However, the illustrations enriched the novel and would be considered important by many readers.

The FRBRization of Humphry Clinker

After completing the broad overview of the *work,* the next step was to identify an *expression* and a *manifestation* for each of the *Humphrey Clinker* bibliographic records. The original, unaugmented *expression* was identified as the "original." The other expressions were named for the editor(s) and/or illustrator(s). When, as occurred once, there were multiple expressions with the same editors, edition numbers were also used. *Manifestations* were named for their publisher and, if necessary, the date of publication. Combining the surnames from the added entries created the initial expression name, with the publisher being used for the manifestation name. For example, the edition edited by Robert Gorham Davis and published by Holt, Rinehart and Winston was identified as the *Davis* expression and the *Holt, Rinehart and Winston* manifestation.

Of the 179 sample records, after excluding the translations and microforms, there were 157

English language print editions. These 157 records were analyzed, and all relevant details for

each record were entered into a spreadsheet. The initial spreadsheet was created by

automatically extracting the relevant information directly from the bibliographic records.

Relevant information included added entries, publisher, pagination, date and place of

publication, statement of responsibility and other similar information. Based on physical

examinations, the spreadsheet was updated to reflect the new observations. Using the filtering

and sorting functions permitted easy clustering of the records using any the attributes. A copy of

the spreadsheet, along with the bibliographic records and the page images, is available on the

project's Web site (OCLC 2002).

All the records were reviewed to correct for insignificant differences in the form of entry, e.g.,

Holt, Rinehart and Winston vs. *Rinehart*. The statements of responsibility were examined to

identify additional editors or illustrators. For example, it was determined that *Robert Gorham*

Davis edited a book only by examining the bibliographic record, which lacked an added entry

but included the statement of responsibility: *Edited with an introd. by Robert Gorham Davis*.

In a separate study, Delsey extensively analyzed the MARC format "to clarify the relationships

between the data structures embodied in the MARC formats and FRBR and AACR models"

(Delsey 2002, 5). He developed a detailed table that associates the various elements in the

MARC record to the attributes of works, expressions, manifestations, and items. In principle,

this table should be able to be used to determine, based on their bibliographic records, whether

two different bibliographic items are members of the same work, expression, or manifestation.

For example, the statement of responsibility (245 field, subfield c) is identified as a manifestation attribute (Delsey 2002). Therefore, if two records have significantly different statements of responsibility, they must represent different manifestations.

The use of Delsey's table was expected to assist in identifying the elements in the MARC record that can distinguish between expressions. To facilitate the use of the table, field and subfield statistics for all 157 English language *Humphry Clinker* records were compiled. Table 1 shows the number of times a field occurred and all of the subfields that were used. For example, the 100 field occurred in all 157 sample records, and the only subfields used were 'a' (Personal name) and 'd' (Dates). The entries in Table 1 were compared to the entries in Delsey's table to identify common elements. The surprising result was that, except for language, there were no common elements. Since none of the expression attributes from Delsey's table occurred in the *Humphry Clinker* bibliographic records, the table could not be used to identify expressions.

When it was difficult to determine if the differences between bibliographic records were real differences or simply differences in cataloging practice, an attempt was made to physically examine one or both of the books. Not all of the books could be obtained since many were either in too poor a physical condition to loan, considered a rare book, or otherwise unavailable for borrowing. In these cases, information was obtained where possible, usually via email directly from one of the holding libraries. It is doubtful that the failure to obtain these books for direct examination had a significant impact on the results, although there may have been a few changes in the assignment of the records to particular expressions.

Results of the analysis are shown in Table 2. The 48 different expressions fell into four distinct

groups: the original, the edited, the illustrated, and the translated expressions. Most of the

expressions were created as the result of an editor adding an introduction, notes and/or a

bibliography; the addition of illustrations; or both. The original expression had 43

manifestations, far more than any of the other expressions. These manifestations were the result

of the expression either being published by a new publisher or being republished with the type

being reset. There were eight translations into seven languages, each with a single manifestation.

Except for these translations, the 39 new expressions were the result of either editors or

illustrators.

The results shown in Table 2 were quite different than the initial version derived solely from the

information in the bibliographic records. In attempts to FRBRize *Humphry Clinker* based only

on the bibliographic records, the most reliable indication that two records represented different

expressions was that their added entries were different. The occurrence of an added entry

indicated that that edition had been edited, translated, or illustrated.

Of the 157 English language records analyzed, 44 had one or more personal name added entries.

An additional 32 edited and/or illustrated records were found by examining the statement of

responsibility, and two more were identified through the notes. Twenty more records were

identified as being edited or illustrated by examining the books themselves. Some illustrators

were identified from their signed illustrations. Many of these signatures, such as Cruikshank's

seen in Figure 4, are brief and can be difficult to read. Overall, 108 of the English language

records represented edited and/or illustrated editions, but only 44 (41%) could be easily

identified from the bibliographic records. Any simple algorithmic approach would incorrectly treat these hard-to-identify expressions as the original expression. More importantly, these unidentified expressions would effectively be lost—undifferentiated from the original expression.

Based on the examination of many of the books and the comparison of a book to its bibliographic description, it became clear that bibliographic records simply do not contain sufficient information to reliably identify expressions. Distinctions based solely on the content of bibliographic records will fail to identify a significant number of expressions and create duplicate expressions based on differing cataloging practice rather than any real differences between the books. For *Humphry Clinker*, expressions identified solely from bibliographic records were unreliable and could impede the navigation process they was designed to assist.

In applying the FRBR entity-relationship model to bibliographic records, the study identified several ambiguities that confounded the FRBRization process. The FRBR report provides an unambiguous definition for *expression* and then proceeds to allow for flexible interpretations. For example, the report states, "if a text is revised or modified, the resulting *expression* is considered to be a new *expression*, no matter how minor the modification may be" (IFLA 1998, 19). Although difficult to implement, this statement is clear and unambiguous. However, in the next paragraph, the report states "On a practical level, the degree to which bibliographic distinctions are made between variant *expressions* of a *work* will depend to some extent on the nature of the work itself, and on the anticipated needs of users." This second statement

contradicts the earlier definition by implying that a standard far more flexible than *no matter how minor* can be employed.

While sufficient flexibility to respond to the needs of various user communities is arguably desirable, the IFLA report does not adequately consider the impact of such flexibility in a shared cataloging environment. In a shared cataloging environment, *consistency* is arguable more important than *flexibility*. While duplicate records can be a problem in any catalog, they are a bigger problem in the shared cataloging environment.

With the FRBR model, the potential for duplicates would exist at three levels— works, expressions, and manifestations. While it would be naïve to assume that duplicates can ever be completely eliminated, the hierarchal FRBR model increases the potential to create large numbers of duplicate records. At the *manifestation* level, duplicates are expected to present similar problems to those currently encountered. However, the problem of duplicate records for manifestations is already serious—over 30% of the *Humphry Clinker* records appear to be duplicates by virtue of their manifestations. By introducing *works* and *expressions*, the FRBR model compounds the duplicate problem. Potentially there can be duplicate records for works that can, in turn, include duplicates records for expressions, which contain duplicate records for manifestations. The problem is further compounded by inconsistent or ambiguous definitions. A large number of duplicate records potentially could limit functionality of the FRBR entity-relationship model.

Are Expressions Valid Entities?

Identifying expressions was problematic and raised the question of whether they are valid

entities. Generally, entities are required to be discrete identifiable objects—not something as

vague as expressions. While some expressions, e.g., translations, are distinct and identifiable,

most of the expressions observed for *Humphry Clinker* were not. Determining if two

manifestations embody the same expression proved to be very difficult. Bibliographic records

rarely contained sufficient information to reliably distinguish expressions, making it frequently

necessary to do either side-by-side comparison or to compare one manifestation to an extensive

set of photographic images of the other manifestation.

Delsey's analysis of the bibliographic format also raises questions as to whether expressions

should be considered entities. None of MARC elements that Delsey identified with expressions

occurred in any of the bibliographic records for the English-language editions of *Humphry*

Clinker. This lack of expression-related elements reinforces the difficulty of using bibliographic

records to identify expressions and helps to explain the difficulties observed.

Is the difficulty of identifying expressions a result of an overly strict definition? Conceptually,

considering any modification to the content *no matter how minor* to result in a new expression

makes sense. The work is a distinct intellectual creation, the expression is the set of all items

with identical content, and the manifestation is a distinct physical unit. In practice, however, it is

extremely difficult to determine if two manifestations have identical content. Even if it could

easily be determined when the content was identical, the result would have an overly fine

granularity—in many cases the distinction between expressions and manifestations would be

lost. New expressions would be created from changes so minor that they would be unnoticed by most readers.

Changing the definition of expression to require that the changes be significant would reduce the problem of trivial expressions but would likely raise other problems. For example, notes, introductions, forewords, bibliographies, and illustrations are significant to some but not all readers. Some contributors may be identified in the statement of responsibility, others may have 'signed' contributions, and others may be completely anonymous. Add the translations to the mix and the difficulty of finding a way to equate the variety of changes becomes very complex, if not impossible. However, unless these changes are equated in a meaningful way, moving beyond the *no matter how minor* standard would be difficult. Building an entity-relationship model that includes expressions may be neither practical nor conceptually sound.

What are the alternatives to expressions? If expressions were dropped from the FRBR model, the model would be greatly simplified but with a significant loss of functionality. There are alternatives that address the same needs which expressions address but are simpler and more responsive to user needs. For *Humphry Clinker*, the increased use of added entries appears to be an effective way to identify expression-like changes. Added entities with the role of the contributor explicitly identified would effectively differentiate among manifestations with different supplemental material. The inclusion of an added entry for all identifiable contributors would require minimal extra effort and, at least for *Humphry Clinker,* would meet the need served by expressions. In effect, expressions could be created dynamically in response to particular user interests.

A reader interested in illustrations could be presented with an expression-like view identifying the illustrator with the number of manifestations illustrated, such as:

1. Allen, Joseph, 1770-1839 (3)

2. Browne, Hablot Knight, 1815-1882 (8)

3. Corbould, Richard, 1757-1831 (4)

4. Cruikshank, George, 1792-1878 (21)

5. Harris, Derrick, 1919-1960 (1)

6. Holloway, Edgar (2)

7. Richards, Frank (8)

8. Rowlandson, Thomas, 1756-1827 (8)

9. Unidentified illustrators (19)

This one-dimensional, illustrator-centric approach presents a clear picture of the illustrators who contributed to *Humphry Clinker* without confounding the bibliographic record with editors, translators, or others contributors. Similar customized views could be constructed for editors and translators.

Replacing expressions with additional manifestation attributes works well in this case for several reasons: It eliminates the difficulty of identifying expressions, it is easier to implement, and it provides the information necessary to dynamically generate custom expression-like bibliographic record displays. For *Humphry Clinker*, replacing the expression in the FRBR model with additional manifestation attributes simplifies the model without any loss of functionality.

Conclusions

The FRBR model provides a powerful means to improve the organization of bibliographic items, particularly for large works such as *Humphry Clinker* where there is no way to navigate easily within the work. *Works* are a valuable concept and provide a means by which to aggregate bibliographic units and simplify database organization and retrieval. It appears that works can be reliably identified from existing bibliographic records. Identifying *expressions*, however, is far more problematic. In the example of *Humphry Clinker*, the set of expressions created from the existing bibliographic records is very different from the set based on the physical examination of the books themselves. The detection of subtle differences, such as an updated bibliography, requires the actual copy of at least one of the books. Existing bibliographic records simply do not contain sufficient information to consistently associate the records with expressions. Attempts to create FRBR expressions from existing records are often futile. If expressions are replaced with manifestation records that included added entries explicitly identifying roles of the contributors, the problem of identifying expressions is avoided without lost of functionality. The remaining entity-relationship structures—works, manifestations, and items—provide a powerful means to improve bibliographic organization and navigation.

The study reported herein developed a data set for a single work, *The Expedition of Humphry Clinker,* and applied the FRBR model to that work. Any conclusions based on a single work are risky and lack statistical justification. However, it is extremely unlikely that the problems encountered with *Humphry Clinker* are unique. Clearly, many of the difficulties are the result of the size of this work—smaller works are likely to present far fewer problems. The irony is that

the FRBR model provides minimal benefits to the small works that can be reliably FRBRized,

but fails on the large and complex works where it is most needed.

References

Anglo-American Cataloguing Rules (1988), 2nd ed., 1988 revision. American Library

Association, Chicago.

Delsey, Tom. Functional Analysis of the MARC 21 Bibliographic and Holdings

Formats, Library of Congress, Network Development and MARC Standards Office,

2002, http://www.loc.gov/marc/marc-functional-analysis/home.html, accessed 3/12/02.

Hickey, Thomas B., Edward T. O'Neill and Jenny Toves. (2002) "Experiments with the IFLA

Functional Requirements for Bibliographic Records (FRBR)," *D-Lib Magazine,* Vol. 8, No. 9,

http://www.dlib.org/dlib/september02/hickey/09hickey.html, accessed 10/18/02.

IFLA Study Group on the Functional Requirements of Bibliographic Records. (1998)

Functional Requirements of Bibliographic Records: Final Report. München: K. G. Saur.

http://www.ifla.org/VII/s13/frbr/frbr.pdf, accessed 3/12/03.

Le Boeuf, Patrick. (2001) "FRBR and Further," *Cataloging & Classification Quarterly*,

Vol. 32, No. 4, p15-52

Lubetzky, Seymour. (1963) "The Function of the Main Entry in the Alphabetical

Catalogue—One Approach," in *Report: International Conference on Cataloguing*

Principles, Paris, 1961. London, Organizing Committee of the International Conference

on Cataloguing Principles, p. 139-143.

OCLC Online Computer Library Center, Inc., Office of Research. *Humphry Clinker Project.*

http://www.oclc.org/research/projects/frbr/clinker

O'Neill, Edward T. and Diane Vizine-Goetz. "Bibliographic relationships: Implications

for the function of the catalog." In E. Svenonius (Ed.), *The conceptual foundations of*

Descriptive Cataloging. p. 167-179. San Diego: Academic Press, 1989.

Smiraglia, Richard P. (2001) *The Nature of "A Work": Implications for the Organization*

of Knowledge. Lanham, Md., Scarecrow Press.

Svenonius, Elaine. (2000) *The Intellectual Foundation of Information Organization.*

Cambridge, Cambridge, MA, The MIT Press.

Tillett, Barbara B. Bibliographic Relationships. (2001) In Carole. A. Bean and Rebecca.

Green (eds.) *Relationships in the Organization of Knowledge*, Kluwer Academic

Publishers, Dordrecht, p 19-35.

Verona, Eva. (1959) "Literary Unit Versus Bibliographical Unit." *Libri* Vol. 9, No. 2, p

79-104.

Verona, Eva. (1963) "The Function of the Main Entry in the Alphabetical Catalogue—a

Second Approach," in *Report: International Conference on Cataloguing Principles,*

Paris, 1961. London, Organizing Committee of the International Conference on

Cataloguing Principles, p. 145-157.

```
┌─────────────────────────┐
│         Work            │
│  A distinct intellectual or │
│     artistic creation   │
└─────────────────────────┘

Is realized
through
                ┌─────────────────────────┐
                │      Expression         │
                │  The intellectual or artistic │
                │    realization of a work │
                └─────────────────────────┘

        Is embodied in
                    ┌─────────────────────────┐
                    │     Manifestation       │
                    │  The physical embodiment of │
                    │   an expression of a work │
                    └─────────────────────────┘

            Is exemplified by
                        ┌─────────────────────────┐
                        │          Item           │
                        │   A single exemplar of  │
                        │     a manifestation     │
                        └─────────────────────────┘
```
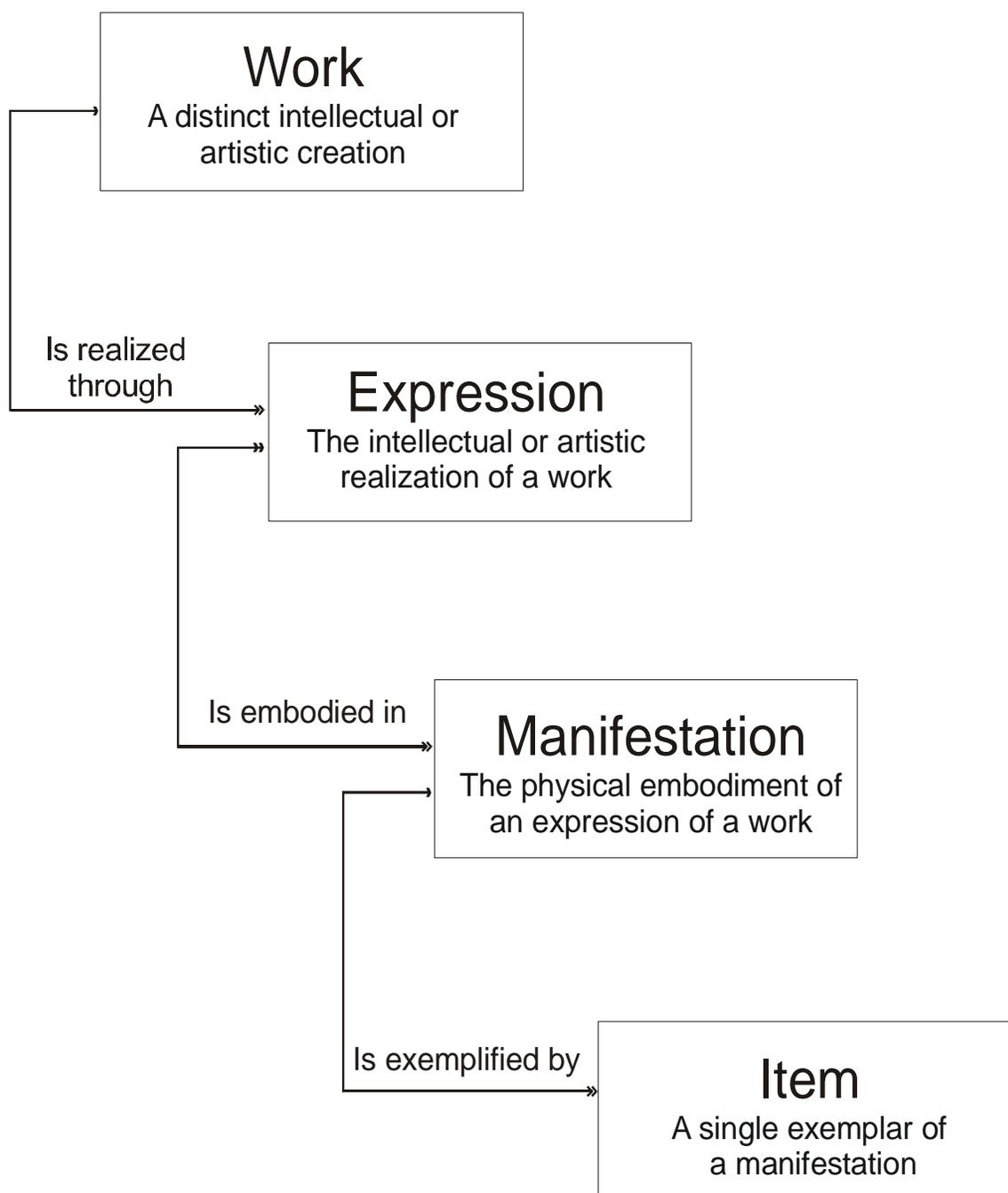
Figure 1. Entities and Primary Relationships

To Mrs GWYLLIM, house-keeper at Brambleton-hall

MRS GWYLLIM,

I can't help thinking it very strange, that I never had an answer to the letter I wrote you some weeks ago from Bath, concerning the sour bear, the gander, and the maids eating butter, which I won't allow to be wasted.—We are now going upon a long journey to the north, whereby I desire you will redouble your care and circumflexion, that the family may be well managed in our absence; for, you know, you must render account, not only to your earthly master, but also to him that is above; and if you are found a good and faithful sarvant, great will be your reward in haven. I hope there will be twenty stun of cheese ready for market by the time I get huom, and as much owl spun, as will make half a dozen pair of blankets; and that the savings of the butter-milk will fetch me a good penny before Martinmass, as the two pigs are to be fed for baking with bitchmast and acrons.

   I wrote to doctor Lews for the same porpuss, but he never had the good manners to take the least notice of my letter; for which reason, I shall never favour him with another, though he beshits me on his bended knees. You will do well to keep a watchful eye over the hind Villiams, who is one of his amissories, and, I believe, no better than he should be at bottom. God forbid that I should lack christian charity; but charity begins at huom, and sure nothing can be a more charitable work than to rid the family of such vermine. I do suppose, that the bindled cow has been had to the parson's bull, that old Moll has had another litter of pigs, and that Dick is become a mighty mouser. Pray order every thing for the best, and be frugal, and keep the maids to their labour.—If I had a private opportunity, I would send them some hymns to sing instead of profane ballads; but, as I can't, they and you must be contented with the prayers of

<div align="center">Your assured friend,</div>

London, June 14.   T. BRAMBLE

Figure 2. Example of a Letter from Humphry Clinker

1972: Recent editions in English include the following: 1929, Modern Library; 1936, Thomas Nelson and Sons, Limited; 1943, Everyman's Library, No. 975, ed. Charles Lee (cited as 'Lee'); 1950, Rinehart and Co.; 1954, in Collins Classics Series; 1955, Folio Society; 1960, Dolphin Books, C. 120; *c.* 1961, Reprint of Everyman Edition, No. 975.

1984: Some recent editions in English: 1925, World's Classics, ed., L. Rice-Oxley; 1929, Modern Library, ed. A. Machen; 1943, Everyman's Library, No. 975, ed. H. M. Jones and C. Lee (cited at 'Lee'); 1950, Rinehart and Co., ed. R. G. Davis; 1954, Collins Classics Series, ed. V. S. Pritchett; 1955, Folio Society; 1960, Signet Classics, C.D. 30; 1966, Oxford English Novels, ed. L. M. Knapp, reprinted 1972 as Oxford University Press paperback, and 1984 by World's Classics (revised and update by P.-G. Boucé); 1967, Penguin English Library, ed. A. Ross (cited as 'Ross'); 1968, Riverside Editions, ed. A. Parreaux (cited as 'Parreaux'); 1968, Heron Books.

1998: Some recent editions in English: 1925, World's Classics, ed., L. Rice-Oxley; 1929, Modern Library, ed. A. Machen; 1943, Everyman's Library, No. 975, ed. H. M. Jones and C. Lee (cited at 'Lee'); 1950, Rinehart and Co., ed. R. G. Davis; 1954, Collins Classics Series, ed. V. S. Pritchett; 1955, Folio Society; 1960, Signet Classics, C.D. 30; 1966, Oxford English Novels, ed. L. M. Knapp, reprinted 1972 as Oxford University Press paperback, and 1984 by World's Classics (revised and update by P.-G. Boucé); 1967, Penguin English Library, ed. A. Ross (cited as 'Ross'); 1968, Riverside Editions, ed. A. Parreaux (cited as 'Parreaux'); 1968, Heron Books; 1983, Norton Critical Editions, ed. J. L. Thorson; 1985, Penguin Classics, ed. A. Ross; forthcoming late 1990, in *The works of Tobias Smollett* (Athens: The University of Georgia Press), ed. T. Preston, the standard and definitive edition; 1991, World's Classics ed., revised and updated by P.-G. Boucé.

Figure 3. Selected sections of the bibliographies from three Oxford University Press editions.

Lismahago's Retaliation.

Figure 4. Example of Cruikshank Illustration.

Table 1.
Fields and Subfields used in
the English Languages Editions

Tag	Field Frequency	Subfields Used
10	22	az
15	12	a
19	7	a
20	15	ac
29	10	ab
35	158	a
40	157	abcde
49	157	a
50	22	ab
82	16	a2
90	85	ab
92	21	ab
100	157	ad
240	4	a
245	157	abcn
246	11	a
250	21	ab
260	157	abc
263	1	a
300	156	abc
440	27	av
490	58	av
500	79	a
504	21	a
510	5	ac
600	2	adtx
650	15	avxyz2
651	12	avxy
653	7	a
655	14	a2
700	62	adepqt45
740	50	a
752	1	abd
800	10	adftv
830	2	a

Table 2. FRBRization Results.

Type of Expression	No. of Expressions	No. of Manifestations	No. of Bibliographic Records	No. of Duplicate Records
Unaugmented	1	43	49	6
Translations	8	8	8	0
Edited	15	24	39	15
Illustrated	13	21	34	13
Edited & Illustrated	11	18	35	17
Totals	**48**	**114**	**165**	**51**