

# FRBR Aggregates

## Their Types and Frequency in Library Collections

Edward O'Neill, Maja Žumer, and Jeffrey Mixter

*Aggregates have been a frequent topic of discussion between library science researchers. This study seeks to better understand aggregates through the analysis of a sample of bibliographic records and review of the cataloging treatment of aggregates. The study focuses on determining how common aggregates are in library collections, what types of aggregates exist, how aggregates are described in bibliographic records, and the criteria for identifying aggregates from the information in bibliographic records. A sample of bibliographic records representing textual resources was taken from OCLC's WorldCat database. More than 20 percent of the sampled records represented aggregates and more works were embodied in aggregates than were embodied in single work manifestations. A variety of issues, including cataloging practices and the varying definitions of aggregates, made it difficult to accurately identify and quantify the presence of aggregates using only the information from bibliographic records.*

Seventeen years after the publication of the Functional Requirements for Bibliographic Records: Final Report (FRBR Report), discussions about the FRBR model have not ceased.<sup>1</sup> Aggregates, which are relatively common, have been a frequent topic of discussion because of their rather vague treatment in the FRBR Report. Aggregates are formed when two or more resources are published together as a unit. Two novels published in a single volume, a book with a foreword, a journal containing many scholarly articles, and a festschrift are examples of aggregates. Varying interpretations of aggregates have surfaced, resulting in a need for clarification.

The approval of the *Final Report of the Working Group on Aggregates* (Aggregates Report) increased the interest in aggregates but failed to resolve all the conceptual issues.<sup>2</sup> This paper approaches these issues from an analytical perspective. Four main research questions are addressed:

1. What types of aggregates exist?
2. How prevalent are aggregates in library catalogs?
3. How are aggregates described in bibliographic records?
4. Can bibliographic records for aggregates be easily identified?

The goal of this study is to gain a better understanding of both the information about aggregate resources recorded in bibliographic records and the nature of aggregates represented by these records.

**Edward O'Neill** (edwardtoneill@gmail.com) is a retired Senior Research Scientist, OCLC, OCLC Research. **Maja Žumer** (maja.zumer@nuk.uni-lj.si) is a Professor at the University of Ljubljana. **Jeffrey Mixter** (mixterj@oclc.org) is a Software Engineer, OCLC Research.

Manuscript submitted July 18, 2014; returned to authors for revision October 6, 2014; revisions submitted December 3, 2014; returned to authors for additional revisions on February 23, 2015; revised manuscript submitted March 8, 2015; accepted for publication March 31, 2015.

## Literature Review

The term *aggregate* is relatively new to the cataloging vocabulary. Neither Chan in *Cataloging and Classification: An Introduction* (2007) nor Taylor in *Introduction to Cataloging and Classification* (2006) include aggregates in their glossaries or indexes.<sup>3</sup> However, neither the concept nor the bibliographic description of aggregates is new. Smiraglia points out that, as early as 1876, Cutter recognized that multiple works could be published in a single physical manifestation and that Cutter “advised distinguishing between the case of joint authors of one work [not an aggregate], and two authors of separate works joined in one volume [an aggregate].”<sup>4</sup>

The FRBR Report provided a new way to think about bibliographic entities and a more precise vocabulary. The four group 1 entities, the products of intellectual or artistic endeavor, are defined as

- *Work*: a distinct intellectual or artistic creation;
- *Expression*: the intellectual or artistic realization of a *work* in the form of alpha-numeric, musical, or choreographic notation, sound, image, object, movement, etc., or any combination of such forms;
- *Manifestation*: the physical embodiment of an *expression* of a *work*; and
- *Item*: a single exemplar of a *manifestation*.<sup>5</sup>

Since aggregates and the group 1 entities are interrelated concepts, any review of the aggregates literature must also include the related FRBR literature.

The FRBR Report was one of the first documents to use the term *aggregate* in the bibliographic context but only briefly discussed aggregates and failed to provide a precise definition. The 2005 workshop *FRBR in 21st Century Catalogues* (FRBR Workshop) is the first known venue that included a detailed discussion of aggregates since the publication of the FRBR Report in 1998. That workshop took place over three days and covered a wide variety of FRBR related topics. The first session was devoted exclusively to aggregates and included presentations by O’Neill, Žumer, Kuhagen, and van Nuys and Albertsen.<sup>6</sup> The presentations covered a variety of issues: the definition of works and aggregates, approaches to modeling aggregates, the treatment of augmentations, continuing resources as aggregates, and the difficulty of retrieving works published as aggregates.

Although the phrasing and details varied considerably, there was general agreement at the FRBR Workshop that multiple works embodied in a single manifestation form an aggregate. However, following the presentations and ensuing discussion, there was no consensus on either the definition or the modeling. This lack of consensus was due, at least in part, to differing concepts of works. Svenonius notes that

as “critical as it is in organizing information, the concept of work has never been satisfactorily defined.”<sup>7</sup> Smiraglia provides a comprehensive review of varying concepts of works beginning with Cutter’s views in the 1870s through the views expressed in the FRBR Report.<sup>8</sup> Without an unambiguous definition of works, the understanding, defining, and modeling of aggregates is problematic.

The publication of the Aggregates Report led to several conceptual papers discussing the report. Žumer and O’Neill’s paper reviewed the manifestation-of-expressions model that was endorsed by the working group; Tillett’s paper described the work-of-works model, which had also been considered by the working group; and Taniguchi’s paper discussed aggregates in the context of RDA.<sup>9</sup>

## Revisions to the FRBR Model

At the FRBR Workshop, expressions and aggregates generated lengthy discussions. Because of the questions raised at the FRBR Workshop, the International Federation of Library Associations and Institutions’ (IFLA) FRBR Review Group established a working group on aggregates. The FRBR Review Group had previously established a working group on expressions.

### Working Group on the Expressions Entity

The Working Group on the Expression Entity was formed in 2003 and tasked “to clarify the *expression* entity and provide application guidelines through examples.”<sup>10</sup> The working group proposed two major changes to the FRBR Report. The FRBR Report stated, “Any change in intellectual or artistic content constitutes a change in *expression*. Thus, 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.”<sup>11</sup> The “no matter how minor” clause proved to be overly strict and resulted in expressions with very minor differences that rarely would be noticed or deemed significant. A detailed comparison of two similar manifestations frequently would disclose some differences, often the result of typesetting errors, spelling differences (colour versus color), or other differences so minor that they would be detected only by a detailed textual comparison.

Recognizing that the “no matter how minor” requirement was impractical and did not serve the users, the working group dropped that requirement and replaced it with “minor changes, such as corrections of spelling and punctuation, etc., may be considered as variations within the same *expression*.”<sup>12</sup> The other change the working group made was to clarify the treatment of augmentations. That change specified that “when an *expression* is accompanied by augmentations, such as illustrations, notes, glosses, etc.

that are not integral to the intellectual or artistic realization of the *work*, such augmentations are considered to be separate *expressions* of their own separate *work(s)*.<sup>13</sup> It was recognized that not all augmentations are significant enough to warrant distinct bibliographic identification. This change created a new type of aggregate; when combined with original text, the supplemental material formed an aggregate whereas previously it resulted in a new expression of the primary work.

### Working Group on Aggregates

The Working Group on Aggregates was formed to evaluate the approaches to modeling aggregates identified during the FRBR Workshop. Specifically, the working group was tasked “to investigate practical solutions to the specific problems encountered in modeling (a) collections, selections, anthologies . . . (b) augmentations, (c) series, (d) journals, (e) integrating resources, (f) multipart monographs, all of which are gathered under the generic term ‘aggregates.’”<sup>14</sup> The fact that the FRBR Report does not clearly distinguish between components and aggregates has been a source of confusion.

To better understand aggregates, the working group collected and discussed numerous examples of aggregates. Various definitions and modeling approaches were applied in an attempt to determine which definitions were most appropriate and which models were consistent with the aggregates examined. While a variety of definitions were explored, the working group focused primarily on two alternative definitions: a broad definition that allowed aggregates to be formed from most FRBR entities based on whole/part relationships and a more limited definition that restricted aggregates to expressions based on the many-to-many relationship between expressions and manifestations shown in figure 3.1 of the FRBR Report. The broader definition would not only allow all four group 1 entities (works, expressions, manifestations, items) to be aggregated but also treat combinations of group 2 (person, corporate body) and group 3 (concept, object, event, place) entities as aggregates. While both definitions have their particular strengths, it was recognized that they were incompatible. While the narrower definition was more restrictive, it was unambiguous and covered all of the resource types that the working group was tasked to investigate. After an extended investigation and discussion, the narrower definition was chosen. An aggregate was defined as “a manifestation embodying multiple distinct expressions” and that aggregates should be modeled as manifestation-of-expressions.<sup>15</sup> This definition does not preclude other groupings based on whole/part relationships but limits the term *aggregate* to manifestations containing two or more expressions.

The essence of the manifestation-of-expressions model is that separate expressions can be embodied in a single

manifestation without creating an encompassing work. When, for example, essays by different authors are published as a collection, it is an aggregate. Each essay is a work, able to stand on its own and is not a component of a larger work. The same may be said for the individual articles contained within a serial issue or volume. It is assumed that structural components, such as chapters of a novel, verses of a poem, or scenes of a play or movie do not form an aggregate; they are identifiable parts of the work and should be modeled as such. A flower is not an aggregate of a stem, leaves, and blossom, and a novel is not an aggregate of chapters.

The *aggregating work* is an important aspect of this model. It is defined as the intellectual contribution of selecting and arranging expressions into an aggregate. The aggregating work has also been figuratively referred to as the glue, binding, or mortar that transforms a set of individual expressions into an aggregation. “This effort may be relatively minor—two existing novels published together—or it may represent a major effort resulting in an aggregate that is significantly more than the sum of its parts (for example an anthology).”<sup>16</sup> It is important to note that the aggregating work does not contain the aggregated works themselves.

The working group went on to identify three types of aggregates: aggregates of collections of expressions (collections), aggregates resulting from augmentation (augmentations), and aggregates of parallel expressions (parallels).

*Collections:* An aggregate is a collection when it consists of expressions of works of the same type. In FRBR, each such work is labeled as independent. Usually an editor or compiler selects and arranges texts, images, or other expressions of works of one or more creators. Anthologies of poetry or short prose, collected and selected works, and scholarly journals comprised of separate articles are typical examples.

*Augmentations:* When an expression is complemented by additional distinct content in a manifestation, such an aggregate is a result of augmentation. In FRBR, this additional content is labeled as “dependent.” Illustrations, forewords, introductions, and biographical essays are the most common examples. The additional content may have a title (for example “illustrations”), but the creator usually differs from the creator of the main work.

*Parallels:* Parallels are manifestations embodying multiple expressions of the same work. They are the easiest to recognize and model. Typical examples include bilingual editions of poetry, multilingual tourist guides, multilingual manuals, etc.<sup>17</sup>

### Effect of Revisions

The revisions made by the two working groups significantly altered the FRBR model. O'Neill's analysis of *The Expedition of Humphry Clinker* provides a practical example of how the changes affected one particular work. *Humphry*

**Table 1.** Expressions Originally Identified

Type of Expression	No. of Expressions
Unaugmented	1
Translations	8
Edited	15
Illustrated	13
Edited and illustrated	11
All expressions	48

*Clinker*, Tobias Smollett's last novel, is a work that has been extensively studied in the FRBR context. It was originally published in 1771 and has been frequently republished. In his 2001 study, O'Neill found 165 bibliographic records and identified 114 distinct manifestations of the work in OCLC's WorldCat.<sup>18</sup> After examining the bibliographic records and, when necessary, an item exemplifying the manifestation, forty-eight distinct expressions were identified. That study did not attempt to identify minor differences, but if the "no matter how minor" criteria had been strictly applied, the number of expressions would have been much higher.

Eight of the *Humphry Clinker* expressions were translations. Excluding the translations, the text of the novel has not changed significantly since it was originally published. All of the other expressions resulted from the addition of introductions, notes, bibliographies, illustrations, and similar augmentations. At least nine different illustrators are known to have contributed to various manifestations. None of the augmentations could be considered integral since many editions are unaugmented and all of the augmented editions were published after Smollett's death. After the Working Group on the Expression Entity amendment, these "illustrations, notes, glosses, etc." became separate works with their own separate expressions. As a result, the number of *Humphry Clinker* expressions dropped from forty-eight to nine: the original English language expression plus the eight translations.

While treating nonintegral augmentations as separate expressions of their own separate works greatly reduced the number of expressions, it created in a new set of works and expressions. The data collected for the *Humphry Clinker* study lacked sufficient detail to reliably estimate the number of expressions and works embodied in each manifestation on the basis of the revised criteria. However, more than thirty different editors and illustrators were identified. Since there were also several unidentified contributors, it is likely that there are many more augmented works. O'Neill's analysis of *Humphry Clinker* revealed forty-eight distinct expressions as shown in table 1.<sup>19</sup> The authors reexamined the original data using the postamendment criteria, and those results are shown in table 2. Before the amendments, all of the

**Table 2.** Revised Frequency and Type of Works

Type of Work	Frequency
Original English expression	1
Nonintegral illustrations	12
Supplemental text	22
All	35

manifestations were considered to embody the same work. After the amendments, *Humphry Clinker* itself is a single work, but now it is estimated that expressions realizing at least thirty-four different works were embodied with expressions of the main work.

O'Neill observed that there was a wide variation in the significance of the supplemental material and that not all warranted bibliographic description.<sup>20</sup> Determining which works are significant is somewhat subjective. Much of the supplemental material, such as a brief dedication, would rarely be considered significant. Illustrations that may not be significant for a literature collection could be significant for an art collection. However, some of the augmentations (introductions, forewords, notes, and illustrations) were extensive and likely to be sought by readers. In many cases, the supplemental materials provided valuable insight into the novel. Presumably these manifestations were acquired largely for their augmentations since there was no obvious need for additional copies of the novel itself.

In estimating the number of works following the revisions, an augmentation was considered significant if its creator was identified in the bibliographic record. There were many manifestations with augmentations for which no editor or illustrator was identified in the bibliographic record. In these cases, it was assumed that the augmentation did not warrant bibliographic identification or description.

## WorldCat Sample

To better understand how to identify and categorize aggregates, the investigators collected and analyzed a sample of bibliographic records. The sample needed to be large enough and varied enough to be statistically significant. It was also important for the bibliographic data to be representative of data commonly held by libraries. OCLC's WorldCat database is the world's largest repository of bibliographic metadata. Additionally, since OCLC's primary partners are libraries, WorldCat data reflects the material typically found in library catalogs. Despite its North American bias, these two factors made the WorldCat database an appropriate source of data for this study.

WorldCat bibliographic records are roughly equivalent to FRBR manifestations, and holding symbols are

roughly equivalent to FRBR items. These equivalences are not exact; not all bibliographic records describe manifestations, and holding symbols may represent multiple items. However, making these equivalences does not introduce a significant bias and is acceptable for the purpose of selecting a representative sample.

Drawing a representative sample from a large union catalog such as WorldCat poses some methodological challenges. As union catalogs grow, the proportion of unique resources grows disproportionately. As of January 2015, WorldCat included 333,518,928 bibliographic records, only a third of which had more than a single holding symbol attached, indicating that they were held by more than one library. However, 91 percent of all holdings symbols were attached to a third of bibliographic records held by multiple libraries. Because of the high proportion of manifestations that are either not held by any library or only held by a single library, a random sample of manifestations from WorldCat would not be reflective of a typical library. Two-thirds of the manifestations in such a sample would be unique resources: archival resources, rare books, manuscripts, and other similar materials. Widely held books from major publishers would be significantly underrepresented in such a sample. Few, if any, libraries have collections with such a high proportion of unique materials and an equally low proportion of commercially published materials.

To overcome this bias inherent in a random sample of bibliographic records from a union catalog, the sample was selected so that the probability of a bibliographic record (manifestation) being selected was proportional to the number of holdings (items) associated with the bibliographic record. A record held by two libraries was twice as likely to be included in the sample as a record held by a single library; a record held by a hundred libraries was a hundred times more likely to be included. This weighting ensured that the resulting sample was representative of the collections of OCLC's member libraries at least in terms of the number of type and uniqueness of the resources.

The sample was restricted to English language textual materials to keep it somewhat homogeneous. Specifically, the bibliographic records in sample were limited to the following:

- English language material (008 Fixed-Length Data Elements, Bytes 35–37 = *eng*)
- English language cataloging (040 field, subfield \$b = *eng*)
- Language materials (Leader Byte 06: Record type = *a* or *t*)

The English language restrictions were pragmatic limitations; a close categorization and analysis of the bibliographic records for non-English materials or for non-English cataloging was beyond the investigators' language skills.

### Reviewing the Sample

Each of the three investigators independently reviewed the sample and coded each entry as an aggregate or a nonaggregate. The investigators also determined the aggregate type and, when appropriate, added a note explaining why they thought it was an aggregate. The analysis was done in multiple steps. After each step, the investigators compared their results and, if necessary, refined the criteria.

The primary question for the investigators was to determine whether the item was an aggregate and, if so, what type. The guidelines in the amended FRBR Report and in the Aggregates Report were initially used to identify aggregates. As anticipated, the existing guidelines were not precise enough, and the investigators frequently failed to agree on the category. When the investigators differed, they met to review the record and attempt to resolve any differences. Often the differences resulted from one investigator missing something or misinterpreting an element, and these differences were quickly resolved. In other cases, the differences resulted from varying interpretation of the guidelines or, more often, conflicting or incomplete data elements in the bibliographic record.

Determining when a particular part or section of a manifestation is a work or simply part of a larger work proved to be particularly difficult. While the Working Group on the Expression Entity amendment introduced the concept of integralness and made it clear that non-integral augmentations are separate expressions, it failed to clearly define the concept. What does it mean to be integral? What are the criteria? In the case of textual augmentations such as introductions, notes, or essays, it was relatively easy to determine whether they were integral. If the authors of the "main" or primary work are different from those of the supplemental texts or the primary text has been published unaugmented or with different augmentations, then the augmentations probably are not integral.

Deciding whether illustrations were integral proved to be more difficult. To address this issue, the investigators used the illustration codes (bytes 18–21 in fixed-length data elements) and the physical description to identify bibliographic records in the sample that represented illustrated manifestations. From this subset, it was determined that many of the bibliographic records in question were associated with children's literature. A second subset consisting only of children's literature was separately reviewed. Children's literature is extensively illustrated, and it was not initially clear how to determine whether the illustrations were integral. The researchers did not initially agree on the integralness of the illustrations of thirty-five illustrated children's books. This review focused on a specific subset of work types and led to an effort to establish a set

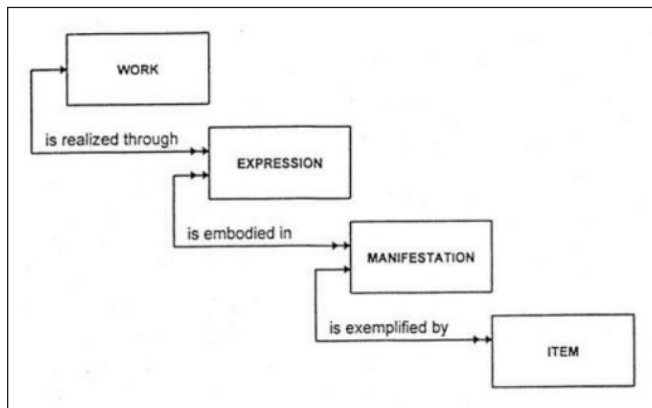


Figure 1. WEMI Hierarchy

of criteria for determining whether illustrations are integral. The issue of integrality, specifically in children's literature, can be difficult to determine in practice because information about the authors' creative process or intent is often unavailable.

After reviewing numerous examples, the authors identified five criteria that were used to determine whether the illustrations are integral to the work:

- the illustrations and the text were created as the result of a collaborative effort
- the illustrations are referred to in the text
- the illustrations were selected by the author for inclusion with the text
- there is a single copyright covering the illustrations and the text
- all known manifestations have the same or similar illustrations

The first criterion acknowledges that many works are the result of a collaborative effort—multiple people working together to create a single work. For example, a physician and a medical illustrator collaborate to create a book on human anatomy. Although the physician and the illustrator played distinct roles, their contributions were coordinated with the intent of creating a single illustrated work. Similarly, many children's books are the result of collaborations between authors and illustrators. In these cases, the illustrations were considered to be integral.

It was frequently difficult to determine whether an illustrated work was the result of a collaborative effort. The second criterion is based on the assumption that if the illustrations are referenced or discussed in the text, they are integral. The FRBR Report is an example of a book meeting the second criterion. As is common with nonfiction works, there are many illustrations in the form of figures, examples, and tables including the frequently

cited group 1 Entities and Primary Relationships in figure 1. The figure is referenced in the text and is an integral part of the work.

The third criterion addresses the case when an author selects preexisting illustrations that were combined with the text. A large number of images are available online from sources such as Getty Images, Flickr, Google Images, and Facebook, plus millions of photographs are available from libraries, historical societies, private collections, etc. An author writing a Cuban travel guide who wants to include a photograph from the Tropicana Club could travel to Havana to take the needed picture. Alternately, the author could select an image from those available online. Considering the wide selection of high-quality images available online, using existing images is a convenient and very attractive option. For historical works, selecting existing images may be the only available option. If the author selects the images, the third criterion is satisfied, and the text and embedded illustrations will form a single work. *Charlevoix's Hotels* (see figure 2) is an example of a book that meets the third criterion. This book provides a historical perspective on hotels in Charlevoix, Michigan, before 1950. The book includes about seventy-five historical photographs, most taken from the Charlevoix Historical Society's collection. The text describing each of the town's hotels is combined with one or more photographs of the hotel. The compilers selected the illustrations and the text and photographs form a single illustrated work.

The fourth criterion concerns copyright assignment associated with the text and illustrations. Copyright assignment provides insight into the relationship between the text and illustrations. If the text and illustrations are separately copyrighted, this could indicate that the illustrations are distinct and are not integral to the text. Copyright can also imply that the illustrations and text were created as part of a collaborative effort. If the text is still under copyright, republishing it with new illustrations would violate copyright unless there was collaboration between the author or the author's agent. Copyright information, particularly as it applies to illustrations, is not consistently included in bibliographic records.

Not all illustrations are integral, and the final criterion, the publication history, can assist in identifying nonintegral illustrations. *Humphry Clinker* is a classic example of a work that has been augmented with nonintegral illustrations. It was originally published without illustrations. Thomas Rowlandson's drawing (see figure 3) was one of the many illustrations subsequently added to augment the main expression. In this case, it is clear that the illustrations were not created as part of a collaborative effort nor did the author select or approve the addition of the illustrations. Therefore the illustrations added to *Humphry Clinker* are separate expressions of separate works.

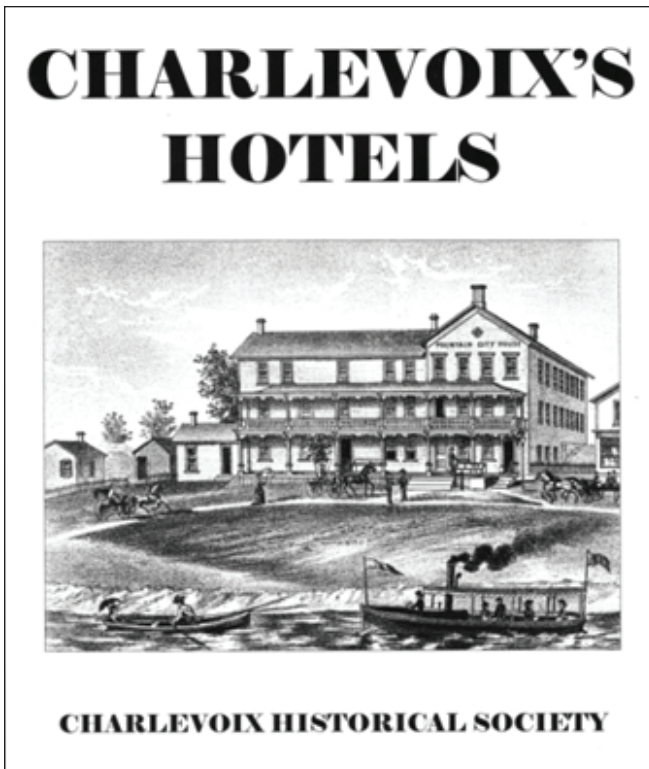


Figure 2. Charlevoix's Hotels

### Limitations of Bibliographic Data

It was often difficult for the investigators to determine whether an item was an aggregate since the decisions were initially based on bibliographic records and descriptions. Even when an item was determined to be an aggregate, it was sometimes difficult to determine the type(s) of aggregate on the basis of the information in the bibliographic record. To resolve these problems, some of the sampled items were obtained from local libraries or borrowed through interlibrary loan. This allowed for a manual review of the item and the associated bibliographic record. When it was difficult to obtain the item and hard to determine whether it was an aggregate, other bibliographic records for the same work were reviewed to see if they contained richer bibliographic information that could be used. When inspecting the actual items, the researchers noticed that they often included additional materials such as introductions, which were not included in the record descriptions. This means that the number of aggregates, as identified from bibliographic records, is underestimated.

One of the primary limitations in determining whether a work was an aggregate was the reliance on bibliographic records. The completeness of the records varied widely,



Figure 3. The Expedition of Humphry Clinker

and it was not uncommon to find records with very limited descriptive information. Conversely, other records included a multitude of additional information that was time consuming to review and process. The disparity between these two extremes highlights the inconsistencies that exist in every catalog but are particularly noticeable in union catalogs. To help overcome this problem during the sample review, efforts were made to identify and use additional bibliographic records that could be used to help determine whether the manifestation was an aggregate. These additional records included duplicate records, parallel records (records cataloged in different languages), and sometimes records that represented different editions of the work. Duplicate records were an easy way to compare two bibliographic records and thus construct a more complete bibliographic description. For these records, the information found could simply be combined. This was particularly beneficial when the primary record (the one included in the sample) was less complete than the duplicate. Records cataloged in different languages were also helpful because, as with duplicate records, they often included information that was not included in the English language record.

The final type of records used to help improve the review process were bibliographic records for different editions of the work. Unlike the information found in the previous types of records, it was not appropriate to combine original bibliographic information with information found in records from different editions. Rather, this information was primarily used to determine whether statements of responsibility for illustrations changed. If so, that was evidence that the illustrations were not integral, and consequently, the work was by definition an aggregate. Even though there were means to overcome the deficiencies found in some of the bibliographic records, the

**Table 3.** Key Attributes

Attribute	Source	Possible Values
Type of resource	Bibliographic level (leader byte 7)	a (monographic component part), i (integrating resource), m (monograph), s (serial)
Number of holdings	WorldCat holdings record	Actual number of holdings
Date of publication	Date1 (bytes 7–10 of 008 field)	Contents of date1
Fiction	Literary form (byte 33 of 008 field)	Yes or no
Juvenile resource	Literary form (byte 33 of 008 field)	Yes or no
Illustrated	Illustrations (bytes 18–21 of 008 field); 245 (c) subfield; 300 (b) subfield;	Yes or no
Brief title	245 (a) subfield	Title proper
Broad subject	Library of Congress Classification	24 broad subject areas derived from the Library of Congress Classification
Genre	Genre	12 general genres
Primary aggregate type	Manual review	Collection, augmentation, parallel
Secondary aggregate type	Manual review	Collection, augmentation, parallel

inconsistencies in cataloging practices posed a problem in judging whether a work was an aggregate.

### Analysis

The investigators analyzed bibliographic records and identified key attributes. The key attributes and their descriptions are included in table 3. Most of the attributes were algorithmically extracted from the bibliographic records and recorded in an Excel spreadsheet. The investigators reviewed and edited the extracted information. When the bibliographic records contained inconsistent or missing data, the attributes were manually assigned. The number of holdings was taken from the WorldCat holdings records. The investigators identified the aggregate types by independently reviewing the bibliographic records and, when necessary, examining an item exemplifying the manifestation or a similar manifestation. The investigators visited libraries, borrowed books, and examined items that were available online. Despite a concerted effort, it was not always possible to determine whether certain manifestations were aggregates. Although it is likely that many of the questionable manifestations are aggregates, these were not categorized as aggregates.

It also became evident that a single manifestation could actually represent multiple types of aggregates. For example, a manifestation containing a collection of short stories with a foreword or introduction is both a collection and an augmentation. When a manifestation was a collection that had also been augmented, it was categorized primarily as a collection.

A total of 212 aggregates were identified from the 1,000 records in the sample. This likely underestimates the actual number because, when in doubt, manifestations were not assumed to be aggregates. Collections were the most common type of aggregate and accounted for 73 percent of the primary aggregates. Collections were frequently augmented with notes, introductions, forewords, and other similar textual material; 23 percent of the collections were also augmented. As a primary type, augmentations accounted for 26 percent of the aggregates, with illustrations being the most common type of augmentation. Parallels were relatively rare, accounting for just 1 percent of the aggregates.

Both the frequency and type of aggregates differed considerably for various resource types. Some types—anthologies and scholarly journals—are, by definition, aggregates. Others—comic books and reference materials—are unlikely to be aggregates. Table 4 lists the major types of the items in the sample. Conference proceedings, scholarly journals, and compilations are similar material types consisting of individual articles or papers, each of which is a separate expression of a separate work. Conference proceedings are often collections of scholarly papers from an academic conference and may be described as either monographs or serials. Scholarly journals are serial publications, and compilations are monographs.

Four types of library materials: anthologies, conference proceedings, scholarly journals, and compilations accounted for almost 15 percent of the sample. These four material types are usually aggregates and can contain thousands of distinct expressions. Considering how common these material types are and the number of expressions they contain, far more works are realized by expressions embodied in



**Table 4.** Aggregates and Material Types

Material Type	No. in Sample	% Aggregates	No. of Collections	No. of Augmentations	No. of Parallels
Anthologies	75	100.0	75	0	0
Archival collections	1	100.0	1	0	0
Children's books	93	8.6	4	4	0
Comic books & graphic novels	4	0.0	0	0	0
Compilations	33	100.0	33	0	0
Conference proceedings	23	100.0	23	0	0
Fiction	98	30.6	0	30	0
Nonfiction	595	4.5	4	21	2
Newspapers & magazines	13	7.7	0	0	1
Reference	39	0.0	0	0	0
Scholarly journals	14	100.0	14	0	0
Transcripts	12	0.0	0	0	0
All material types	1,000	21.2	154	55	3

aggregates than are realized in expressions embodied in nonaggregate manifestations. Fifty-eight percent of the manifestations in the sample were illustrated, but in all but eight cases, the illustrations were considered to be integral.

## Conclusion

Aggregates are very common; more than 20 percent of the library resources sampled were aggregates. While some aggregates embody only a few expressions, scholarly journals, conference proceedings, and compilations usually embody a large number of expressions. More works are realized by expressions embodied in aggregates than in nonaggregate manifestations. Despite their frequency, aggregates are not well understood, lack an accepted definition, and are cataloged inconsistently.

This study confirmed the findings of the Working Group on Aggregates, at least for textual materials. Only three distinct types of aggregates were found: collections, augmentations, and parallels. This study went beyond the working group report by estimating that 73 percent of the aggregates were collections, 26 percent were augmentations, and 1 percent were parallels. It was also observed that these three types of aggregates were not mutually exclusive. In particular, a significant number of collections were also augmented.

Reliably identifying aggregates based on the information in bibliographic records proved to be problematic. The two working groups established by the FRBR Review Group clarified many of the issues associated with expressions and aggregates. However, even utilizing the findings of these

working groups, it was found that bibliographic records frequently lack the detail necessary to determine whether the manifestation described is an aggregate; that is, whether the manifestation embodies multiple expressions.

The FRBR Working Group on the Expression Entity introduced the concept of integrality by stating that augmentations that are not integral to the intellectual or artistic realization of the work are separate expressions of their own separate work. While the guidelines provided by the working group for determining the integrality were generally adequate for textual augmentations, they did not provide sufficient guidance for illustrations. To assist in determining the integrality of illustrations, the authors proposed five criteria.

The ease of identifying aggregates varied considerably. Collections and parallels were generally easy to identify, but augmentations were more problematic primarily because of the difficulty of determining whether the supplemental material, particularly illustrations, was integral to the work. Evaluating the significance of augmentations is highly subjective, and their significance varies widely. Many augmentations were not considered significant enough by the cataloger to be explicitly identified or described and were discovered only when the publications were examined. Some, such as a simple dedication, rarely warrant bibliographic description. Others, such as extensive notes or comments, are often sought by readers and do warrant bibliographic description.

## References

1. IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic*

- Records: Final Report* (München: K.G. Saur, 1998).
2. IFLA Working Group on Aggregates, *Final Report of the Working Group on Aggregates* (September 12, 2011), accessed November 25, 2014, [www.ifla.org/node/923](http://www.ifla.org/node/923).
  3. Lois Mai Chan and Theodora Hodges, *Cataloging and Classification: An Introduction* (Lanham, MD: Scarecrow, 2007); Arlene G. Taylor and David P. Miller, *Introduction to Cataloging and Classification* (Westport, CT: Libraries Unlimited, 2006).
  4. Richard P. Smiraglia, *The Nature of "A Work": Implications for the Organization of Knowledge* (Lanham, MD: Scarecrow, 2001), 19.
  5. IFLA Study Group, *Functional Requirements for Bibliographic Records*, 16–23.
  6. FRBR in 21st Century Catalogues: An Invitational Workshop, "Program and Proceedings" (Dublin, OH, May 2–4, 2005), accessed October 7, 2014, [www.oclc.org/research/activities/frbr/frbr-workshop/program.html](http://www.oclc.org/research/activities/frbr/frbr-workshop/program.html).
  7. Elaine Svenonius, *The Intellectual Foundation of Information Organization* (Cambridge, MA: MIT Press, 2000), 35.
  8. Smiraglia, *Nature of "A Work."*
  9. Maja Žumer and Edward T. O'Neill, "Modeling Aggregates in FRBR," *Cataloging & Classification Quarterly* 50, no. 5–7 (2012): 456–72; Barbara B. Tillett et al., "Letter to the Editor," *Cataloging & Classification Quarterly* 52, no. 3 (2014): 362–64; Shoichi Taniguchi, "Aggregate and Component Entities in RDA: Model and Description," *Cataloging & Classification Quarterly* 51, no. 5 (2013): 580–99.
  10. "Cataloguing Section, FRBR Review Group, Working Group on the Expression Entity," IFLA, accessed November 25, 2014, [http://archive.ifla.org/VII/s13/wgfrbr/expression\\_wg.htm](http://archive.ifla.org/VII/s13/wgfrbr/expression_wg.htm).
  11. IFLA Study Group, *Functional Requirements for Bibliographic Records*, 19.
  12. IFLA Study Group on the Functional Requirements for Bibliographic Records, *Functional Requirements for Bibliographic Records: Final Report*, amended and corrected (The Hague: International Federation of Library Associations and Institutions, 2009), 19, accessed March 3, 2015, [www.ifla.org/VII/s13/frbr/frbr\\_2008](http://www.ifla.org/VII/s13/frbr/frbr_2008).
  13. *Ibid.*, 20.
  14. "FRBR Review Group, Working Group on Aggregates," IFLA, accessed October 7, 2014, [www.ifla.org/node/923](http://www.ifla.org/node/923).
  15. IFLA Working Group on Aggregates, *Final Report*, 3.
  16. *Ibid.*, 5.
  17. *Ibid.*, 3–4.
  18. Edward T. O'Neill, "FRBR: Functional Requirements for Bibliographic Records: Application of the Entity-Relationship Model to *Humphry Clinker*," *Library Resources & Technical Services* 46, no. 4 (2002): 158, [dx.doi.org/10.5860/lrts.46n4.150](https://doi.org/10.5860/lrts.46n4.150).
  19. *Ibid.*, 156.
  20. *Ibid.*, 154.