Multiple-Version Resources in Digital Libraries: Towards User-Centered Displays

Zorana Ercegovac
Henry Samueli School of Engineering and Applied Sciences, 4731H Boelter Hall, University of California–Los Angeles, 405 Hilgard Avenue, Los Angeles, CA 90095. E-mail: zercegov@ucla.edu

The author reports findings from experiments with the International Federation of Library Associations and Institutions' (IFLA) Functional Requirements for Bibliographic Records (FRBR) as applied to the domain of science fiction, Edwin A. Abbott's Flatland: A Romance of Many Dimensions in the Online Computer Library Center's (OCLC) WorldCat. The goal of the study is to gauge the characteristics of bibliographic entities under study, to examine types of relationships these entities exhibit, and to collocate bibliographic entities according to the FRBR group 1 hierarchy of entities identified as works, expressions, manifestations, and items. The study's findings show that by assembling bibliographic records into interrelated clusters and displaying these according to the FRBR entity-relationship model, a new navigational capability in networked digital libraries can be developed.

Introduction

Digital libraries (DL) are systems providing a community of users with coherent access to a large, organized repository of information and knowledge (Lynch & Garcia-Molina, 1995). Coherent access to DL collections hinges on the organization of resources in a collection. Organization is influenced by the nature of resources and patterns of users' tasks as well as by many other factors, including technology, protocols, and policy. The content of resources has increasingly become dynamic, ranging from minor updates, to replacements and multiple releases. Contents may be represented in different forms of expression (e.g., textual, numeric, graphical, video, and multimedia); new genres or classes of materials used in various formats (e.g., html, postscript, gif, tiff, jpg, txt); and encoded with mixed metadata (e.g., MARC, XMLMARC, XML, Dublin Core, Resource Description Framework, RDF, RFC 1807, automatically generated). Digital libraries, as defined above, use disparate architectures and metadata to describe resources, ranging from loosely federated resources on the Web that are searchable by any of the available search engines, to well-structured online library catalogs (Ercegovac, 1999). While the latter produces a display of more consistently described entities than the former, there is room for improvement, especially in the areas of linking entities together and displaying them in an order that would facilitate the decisions of users to select and ultimately obtain desired manifestation(s) of a single work from a collection. The FRBR study offers entity-relationship analysis that may well be the blueprint for knowledge organization in global digital libraries (GDL). Currently, it is the manifestations [§ 2.2B in the AACR2R (Anglo-American Cataloguing Rules; American Library Association [ALA], 2002) such as “2nd ed.,” “New ed.,” “rev. and enl.,” “1st American ed.,” etc.] that can be obtained from a collection. The term manifestation is used to represent a set of all the physical objects that bear the same intellectual or artistic content and physical features of a given work.

The quality of digital libraries depends on the ability to link entities through bibliographic relationships; to access entities (through names, words, or phrases in a record that may be used to find that record in a retrieval system); and to describe entities (through attributes such as a “creator” and title of a work, edition information, publication and distribution data, series, notes, subjects, availability, and intellectual rights management). Conventional libraries have emphasized the description part; global digital libraries will most likely focus their attention on the ways these resources may be accessed, linked, and displayed. Because collections in GDLs are to be shared and used internationally, the GDL will ultimately become more interoperable, cooperative, standardized, and integrated than ever before, which is the goal of universal bibliographical control (UBC). A critical aspect of UBC is the work on navigational pathways or links between and among entities (Ercegovac, 2000; Library of Congress, 2003; Tillett, 1987, 1992a, 1992b, 1997, 2001). This work was started while considering the group 1 entities (work, expression, manifestation, and item) as defined in
The following studies have contributed in major ways to our understanding of the interrelated concepts of multiple versions, defined as alternative manifestations of a single work, and of bibliographic relationships, defined as an association between at least two entities, or components of entities. A bibliographic family is defined as “a set of related bibliographic works that are somehow derived from a common progenitor” (Smiraglia & Leazer, 1999, p. 494). A series of studies on linking bibliographic entities are reviewed and discussed in Smiraglia (2002, pp. 1–11). These include IFLA’s definitions in Functional Requirements for Bibliographic Records (1998); research on “superworks” by Carlyle (1999), Yee (1998), and Svenonius (2000); a proposal by Leazer and Furner for the graphic representation of “textual identity networks” (1999); and Smiraglia’s definitions of “works” (2002, pp. 1–11). He broadened the scope of the term bibliographic families and adopted the term instantiation network. The newly coined definition retains the concept of connectedness, moves beyond the textual, and indicates “a sense of temporality” (p. 7).

Taxonomy of Bibliographic Relationships

The problem of multiple versions goes back to cataloging practice during the pre-electronic era (Lubetzky, 1969; Verona, 1959; Wilson, 1968).

In the electronic environment, the following studies recognized these types of bibliographic relationships: the UNIMARC format (IFLA, 1980) recognized vertical (hierarchical), horizontal relationships between different versions of work in different languages, formats, media, and with chronological relationships. Graham (1990) reviewed proposals to deal with the crisis of multiple versions in cataloging in view of the overall goals to improve access, streamline cataloging processes, and reduce costs. She directed attention to equivalence relationships, as defined by Tillett (2001), and reviewed version-specific elements that are indicative of equivalence relationships (e.g., MARC fields 007, 008, 245 subfield $h, 260 subfield $c, and 300). Tillett (1987, 1992a) found that between 11.2% and 19.4% of Library of Congress (LC) catalog records exhibit derivative bibliographic relationships. Drawing on the content of library cards from the Georgetown University Library (Washington, DC), Smiraglia (1992, p. 60) determined that 49.9% of the works in his sample exhibited a derivative bibliographic relationship. Vellucci (1995) examined the relationships found among musical bibliographic entities that exist in library collections. She discovered that 97% of the scores in her sample drawn from the catalog of the Sibley Music Library (Eastman School of Music, University of Rochester, Rochester, NY) exhibited at least one derivative relationship.

Smiraglia and Leazer (1999) determined the proportion of bibliographic families, more recently termed instantiation network, exhibiting derivative relationships to be 30.2% in a sample of 477 progenitor works in the Online Computer Library Center’s (OCLC) WorldCat (p. 494). They found that popularity of derivative works is greater in academic and research library collections than it is in online union catalogs such as WorldCat, and that popularity influences the size of a bibliographic family. Drawing on a random sample of 968 bibliographic records from the WorldCat, Bennett, Lavoie, and O’Neill (2003) found that less than 25% of works had multiple manifestations. While reviewed results vary in distributions of established derivative bibliographic relationships, researchers agree that bibliographic relationships are prevalent and should be accounted for in designing online library catalogs.

1. Simultaneous derivations (works that are published in two editions simultaneously, or nearly so)
2. Successive derivations (revisions of the same work one or more times, and labeled as numbered editions)
3. Translations
4. Amplifications (with illustrations, commentaries)
5. Extractions (abridgements, condensations)
6. Adaptations (simplifications; screenplays)
7. Performances (sound or visual recordings)
Multiple-Version Resources in Global Digital Libraries

Researchers have pointed out that “the information that was sufficient to identify particular items and bibliographic relationships within a local collection is inadequate in this new environment” (Ercegovac, 1998; O’Neill & Vizine-Goetz, 1989, p. 172). Graham (1990) reviewed proposals to deal with the crisis of multiple versions in cataloging in view of the overall goals to improve access, streamline cataloging processes, and reduce costs. Recently, the Library of Congress (2002) has acknowledged, “Within the last few years, the issue has become even more pressing as a result of the increasing use of digital technologies both to create original materials and to convert existing materials to digital form, and the resulting proliferation of alternative manifestations of the same content.” The Library’s Network Development and MARC Standards Office (NDMSO) illustrated how a hierarchical display of records for multiple manifestations of the same work could be generated from data contained in separate MARC21 records. The NDMSO experimented with Sister Carrie by Theodore Dreiser (http://www.loc.gov/marc/marc-functional-analysis/multiple-versions.html#marc-sister) and The English Patient by Michael Ondaatje (http://www.loc.gov/marc/marc-functional-analysis/multiple-versions.html #marc-english).

Recently, researchers have experimented with algorithms that group existing bibliographic records of a general nature into works and expressions (Hickey, O’Neill, & Toves, 2000). Others have developed models that facilitate interoperability between metadata ontologies from different domains (Lagoze & Hunter, 2001). The emergence of the Web has had the effect of producing an abundance of multiversioned XML resources across disciplinary research. This has also been discussed in computer science literature under various headings such as clustering, knowledge representation, historical or temporal XML documents, time series algorithms, query support, mining streams, and conceptual modeling (Association for Computing Machinery [ACM] Special Interest Group on Management of Data [SIGMOD], n.d.; Chien, Tsotras, & Zaniolo, 2002).

The Functional Requirements for Bibliographic Records’ Entity-Relationship Model: The Basis for the Next Generation of Knowledge Organization in Global Digital Libraries

According to the FRBR (FRBR http://www.ifla.org/VII/s13/frbr/frbr.pdf), there are three groups of entities: Group 1 includes works, expressions, manifestations, and items; group 2 includes persons and corporate bodies that are responsible for intellectual and/or artistic content, while group 3 pertains to subjects of works such as concepts, objects, events, and places.

Accordingly, work is a distinct intellectual or artistic creation. Examples of a work as an abstract entity are Shakespeare’s Romeo and Juliet, Berlioz’ Symphonie Fantastique, Edwin A. Abbott’s Flatland: A Romance of Many Dimensions, etc. Works that are substantially modified through the medium of expression become new works (per AACR2R rule 21.9). Specifically, catalogers “enter a paraphrase, rewriting, adaptation for children, version in a different literary form (novelization, dramatization) under the heading for the adapter” (AACR2R rule 21.10A). In the database under study of 86 records, Abbott’s Flatland led to four new works such as a motion picture and sound recordings.

The FRBR considers expressions as 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 forms. An example of an expression of the Berlioz’s Symphonie Fantastique is a sound recording (1993) by the Paris Symphony Orchestra with Pierre Monteux conducting; another expression is a sound recording (1967) by the Toronto Symphony with Seiji Ozawa conducting. A new expression is any change in language (AACR2R, Rule 21.14A), revision and/or edition (Rule 21.12A1), no matter how minor the modification may be.

Manifestation is defined as the physical embodiment of an expression of a work, encompassing a wide range of materials, such as books, periodicals, maps, posters, sound recordings, films, video recordings, CD-ROMs, and multimedia kits. Manifestation represents all the physical objects that bear the same intellectual and artistic characteristics, both in content and physical form. The physical embodiment of a given expression of a work may be paper or a computer file as well as text, a sound recording, visual material such as canvas, plaster, clay, or other material. In some cases, there may be a single physical exemplar (e.g., Van Gogh’s painting, Irises; Abbott’s first edition Flatland of 1884). In other cases, multiple copies are reproduced to facilitate commercial and educational uses (e.g., irises on note cards, on posters, in art history books). Each manifestation necessarily has the publisher’s layout and typesetting, color, size, packaging, and other details. An example of a hierarchical display of Symphonie Fantastique might be as follows:

Work: Berlioz’s Symphonie Fantastique

- Expression #1: [Sound Recording] 1967, Toronto Symphony, Seiji Ozawa conducting.
- Manifestation #1: New York: CBS Records, 1 sound disc: 33 1/3 rpm, mono.; 12 in.

When catalogers identify and describe an object in hand, they do so at the level of manifestation and generalize to all exemplars or copies that might be located on different library shelves. So an item is an exemplar of the manifestation, such
as an autographed copy, a copy having notes in margins, an item with missing pages, or the one that is located in a particular library.

The Entity-Relationship (E-R) model enables us to draw relationships between a work and related works, between a work and its expressions, and so on. Hence, we can relate the work of Abbott’s *Flatland* to the work of criticism and interpretations, to various expressions (translations into Italian or Hebrew), and to new works such as a film or a sound recording based on the parent work. Tillett’s taxonomy and Smiraglia’s modification were used as a starting point for identifying relationships among entities.

Research Questions

The following questions were considered from the outset of this project:

1. What types of science fiction *entities* exist in collections under study? What categories of *bibliographic relationships* are found in a sample of entities in a given collection?
2. What is the capability of expressing bibliographic entities and relationships in current cataloging standards?

One of the objectives of information retrieval systems, library catalogs included, is to help the user understand the options and content of the database the user is searching. User studies have pointed out that end users have serious problems when they search online library catalogs with regard to scanning long displays (e.g., Matthews, Lawrence, & Ferguson, 1983; Wiberley, Daugherty, & Danowski, 1995). However, research from multiple studies suggests that display of bibliographic records in current online public access catalogs (OPACs) if reorganized differently, might be more helpful to end users. For example, authors who studied bibliographic relationships (e.g., Bennett, Lavoie, & O’Neill, 2003; Smiraglia, 1992, 1994; Smiraglia & Leazer, 1999; Tillett, 2001; Vellucci, 1995) have demonstrated that derivative relationships are important in library collections. If the prevalence of bibliographic families, defined as a set of related works that are derived from a common original resource, is accounted for in the design of library catalogs, assembling entities into clusters could be helpful to users in their decision-making process. Other studies have suggested ways to improve bibliographic displays by means of clustering entries together (e.g., Buckland, Norgard, & Plaunt, 1993; Carlyle, 1996, 1999, 2001; Carlyle & Summerlin, 2002). Yet other sets of studies on human cognitive information processing (e.g., Card, Newell, & Moran, 1983; Miller, 1956) remind us that human capacity to receive, process, and remember information is finite and limited to about seven units of information.

Therefore, we need to take advantage of distributions that seem to be operating among derivative bibliographic relationships. When applied, the E-R model would have the effect of collocating like entities together and displaying them in a way that is native to human information processing capacity, which, in turn, would better facilitate searchers’ decision-making processes. Specifically, the model would collocate different manifestations and expressions of the same work, and by extension, show new works that are based on the common original publication. Expressive linkages between entities would enhance the navigational capability in terms of enabling us to group related entities together and show how these entities are collocated. By showing explicit relationships among groups of records, the model would help the user navigate search results and facilitate tasks of resource discovery, use, and data management.

Design of different views should support various communities of users for their specific tasks and purposes. Such views could be designed to help the public to navigate through their searches, to guide school students, and to provide direction to professionals involved in creating, illustrating, translating, publishing, reviewing, cataloging, archiving, preserving, and intellectual property rights management. Each of these groups has different priorities, capabilities, and preferences that should be accommodated in the user interface design.

Displaying Bibliographic Entities

The model presented in Table 1 provides the general framework of this exercise. The model based on a sample of 86 records shows various search options that the user might explore in the process of his or her search. A typical search scenario would serve the novice searcher who wishes to get an overview of the information space for Abbott’s *Flatland*.

Accordingly, an author/title query would display a hierarchical view of *Flatland*. At the top, the user identifies Abbott’s *Flatland* as the desired entity, and observes Work 1 with 79 records. As Table 2 illustrates, all English-language printed manifestations are grouped into expressions and all are based on the original publication, the English text published by Seeley in 1884 (record #21).

The user also notices new works that are based on Abbott’s *Flatland*. As Table 3 illustrates, these include video recordings, animations, and a motion picture. If one chose the nonbook entities (works 2–5), and selected the animated film group (W4), the user is then presented entries in a layered manner. First, she or he sees three releases or manifestations of video recordings as follows: a black and white release produced by Media Magic in 1994, another manifestation in color produced by Great Media in 1994, and the

<table>
<thead>
<tr>
<th>TABLE 1. Distribution of Abbott’s <em>Flatland</em> under study.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abbott <em>Flatland</em> is a distinct intellectual creation.</td>
</tr>
<tr>
<td><strong>WORK 1: Textual entities</strong></td>
</tr>
<tr>
<td>—English language books ( n = 70 )</td>
</tr>
<tr>
<td>(see Tables 2, 4–7 for detail)</td>
</tr>
<tr>
<td>—Non-English language books ( n = 9 )</td>
</tr>
<tr>
<td><strong>WORKS 2–5: Non-textual entities</strong></td>
</tr>
<tr>
<td>(see Table 3 for details)</td>
</tr>
<tr>
<td>Total entities in collection under study ( N = 86 )</td>
</tr>
</tbody>
</table>
There are serious issues with the newly created display of “frbr-ized” records. Mainly, the display depends on current prevalence of bibliographic families in collections and from the human information processing research. Current displays lack (a) the capability to group manifestations into expressions, (b) display those expressions under a given work, and (c) alert users of other related works. We need to exploit available information from MARC records—much of which is hidden from the user—yet is potentially useful in decision making during browsing and searching.

**Method**

This study’s starting points were definitions for Group 1 entities (i.e., work, expression, manifestation, item), products of intellectual and artistic endeavor, as reported in the FRBR study, and Tillett’s taxonomy of content relationships between and among entities (1992a, 2001) with Smiraglia’s modification (1992). A descriptive survey methodology was used to examine cataloging entries of records from UCLA’s online library catalog Orion, University of California—Melvyl, Library of Congress catalog, and the OCLC’s WorldCat. The Orion library catalog was selected to allow for physical examination of local resources. The breakdown of sample records found for Abbott’s *Flatland* science fiction text (1884) is shown below.

<table>
<thead>
<tr>
<th>Source</th>
<th>Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>WorldCat (OCLC union catalog)</td>
<td>86</td>
</tr>
<tr>
<td>Orion (UCLA libraries)</td>
<td>10</td>
</tr>
<tr>
<td>Melvyl (University of CA)</td>
<td>19</td>
</tr>
<tr>
<td>Library of Congress catalog</td>
<td>13</td>
</tr>
</tbody>
</table>

All 86 records for Abbott’s *Flatland* were searched exhaustively in March 2003, and again in April 2004. An author-title search ([au: abbott and au: edwin] and ti: flatland) for Abbott’s *Flatland* produced a total of 89 bibliographic records in WorldCat. Two duplicate records were discarded from the final data set, and one record was incomplete for the purposes of this study. It is likely that some translations were missed since the use of uniform title is optional (rule 25.1., AACR2R). Both Vellucci (1995) and Carlyle (1997) offer an in-depth discussion of the classificatory function of uniform titles.

Next, the FRBR Display Tool (http://www.loc.gov/marc/marc-functional-analysis/tool.html) was applied on MARC21 *Flatland* records that were retrieved from the Library of Congress catalog (fewer in number than from the set of previously obtained WorldCat records). The most recent search on the newly installed library catalog at UCLA (http://catalog.library.ucla.edu/webvoy.htm) yielded again 10 entries (as of July 15, 2004). Because the Library of Congress search set was slightly larger, it was used with the FRBR display tool. It “sorts and arranges bibliographic record sets using the FRBR model. It then generates useful hierarchical displays of these record sets containing works that consist of multiple expressions and manifestations” (Library of Congress NDMSO, 2004).

There are serious issues with the newly created display of “frbr-ized” records. Mainly, the display depends on current...
cataloging rules (AACR2R; ALA, 2002) and the existing MARC format, neither of which is based on the FRBR E-R model. How to collocate expressions and display together all such realizations of a work that is in a given database is not simply an interface question.

The present study is just a minute step in the effort to transform current lists of cataloging entries into hierarchical displays. The approach has been the combination of the existing links for collocating entities into groups, and inducing it by discovering patterns of text strings that are indicative of collocating entities into groups. A complementary approach is to use systems power during searching and sorting results.

A data set of 86 records was examined manually and, for future experimentation, exported into a FileMaker® program. Records were coded with MARC tags that are normally used in current cataloging practice. New fields were added to facilitate this project’s analysis, matching, sorting, and reporting. These include the field titled “parent,” separately tagged subfields in the statement of responsibility (245 $c1, $c 2,..., $n 3) to recognize various authorial responsibilities (e.g., illustrator, writer of prefaces or introductory notes, a bibliographer who writes annotations and prepares references), the “original publication,” and the “expression_type” field. The “original expression” in the case of Abbott’s Flatland, was published in London by Seeley in 1884 as English text. It has no parent. These newly added fields have facilitated the analysis in this project because the FRBR display tool, for expression level, matched Leader/06 and field 008/35–37 for the language of expression.

In addition to different language expressions, we wanted to create clusters based on derivative bibliographic relationships, including amplified and successive or numbered editions, the ones that are “new and revised,” and special editions (e.g., by name or physical characteristics).

The field parent is used to code those entities that produced children. The newly created parent field in this study made it possible to study clustering of entities within the bibliographic genealogy of Flatland. Findings may shed light on ways to change distances between nodes dynamically.

The field expression_type is defined as a holder for specific types of expression that were found in the dataset under study (e.g., amplified edition, amplified and numbered edition, translated edition, illustrated). Tags for content relationships were also added, including those for equivalence, for descriptive, shared, and whole–part relationships.

Results

The Big Picture

By means of including the parent field, we could see the entire landscape of the Flatland genealogy. For example, the 1884 expression (record #21) is listed as a parent record; it parented four prolific editions represented in records #22, #28, #41, and #44 (Table 2).

Record #22 produced two siblings, one of which (record #24) parented four editions of which record #30 produced eight new manifestations, all by the same publisher (Little, Brown, Boston, MA) between 1927 and 1939 with minor variations in the physical description area. As shown in Tables 4–7 below, four records (#22, #28, #41, and #44) populated the entire bibliographic space consisting of English-language book entities.

Of the 86 retrieved records on Abbott’s Flatland, all translations (n = 9) were based on the original publication. All new works (n = 7) were also based on the 1884 publication. Of the remaining 70 records, 36 came directly from the four major offsprings; this is 57% (40/70) of all English-language books in the sample. This sort of distribution falls

<table>
<thead>
<tr>
<th>TABLE 4. Expressions and manifestations produced by #22.</th>
</tr>
</thead>
<tbody>
<tr>
<td>#22 (n = 15) Boston: Little, Brown</td>
</tr>
<tr>
<td>#23 Boston: Roberts Brothers, 1891 155p. ill 18 cm.</td>
</tr>
<tr>
<td>#24 Boston: Little, Brown and Co, 1899. Manifestations:</td>
</tr>
<tr>
<td>#25–#27 (1907, 1912, 1915)</td>
</tr>
<tr>
<td>#30 is new expression (1926 “with intro. by William Garnett.”)</td>
</tr>
<tr>
<td>It gives 8 manifestations, #31–4; 36–9, 1927–1941.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 5. Nine expressions produced by #28.</th>
</tr>
</thead>
<tbody>
<tr>
<td>#28 (n = 10) Oxford: Blackwell</td>
</tr>
<tr>
<td>(3rd ed., rev. with intro by William Garnett, 1926)</td>
</tr>
<tr>
<td>#29 3rd ed., rev. 1926, intro by W. Garnett. (incomp. cataloging)</td>
</tr>
<tr>
<td>#40 5th ed., rev. 1944, with intro by W. Garnett.</td>
</tr>
<tr>
<td>#42 5th ed. rev. 1950, with intro by W. Garnett.</td>
</tr>
<tr>
<td>#43 6th ed. rev. 1950, intro . . .</td>
</tr>
<tr>
<td>#48 1962, intro . . .</td>
</tr>
<tr>
<td>#47 1962, intro . . . ( #47 and #48 may be the same entity)</td>
</tr>
<tr>
<td>#54 2nd rev. ed., 1874, with intro by W. Garnett.</td>
</tr>
<tr>
<td>#56 1978, with intro . . .</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 6. Expressions produced by #41.</th>
</tr>
</thead>
<tbody>
<tr>
<td>#41 (n = 8) New York: Barnes &amp; Noble</td>
</tr>
<tr>
<td>#52 . . . with intro by William Garnett, 1964, “Followed by Dionys Burger’s Sphereland.” (shared relationship by topic)</td>
</tr>
<tr>
<td>#4 5th ed., rev., 1969 (incomplete cataloging)</td>
</tr>
<tr>
<td>#63 (reprint of: 5th ed., rev., 1963, with foreword by Isaac Asimov); with intro by William Garnett; series Everyday handbook; EH/573.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 7. Expressions produced by #44.</th>
</tr>
</thead>
<tbody>
<tr>
<td>#44 distribution (n = 5) New York: Dover</td>
</tr>
<tr>
<td>(6th ed., rev. with introduction by B. Hoffmann, 1952)</td>
</tr>
<tr>
<td>#2, 6th ed., rev. with intro by B.H., 1952</td>
</tr>
<tr>
<td>#45 6th ed. re., with intro by B.H., 1953</td>
</tr>
<tr>
<td>#46 7th ed., rev. with intro by B.H., 1957</td>
</tr>
</tbody>
</table>
within Bradford’s distribution of various information units, e.g., books, catalog entries, and subject headings (Bates, 2003).

**The Flatland Family Displayed in Current Online Public Access Catalogs**

The work, identified by author and title (Abbott/Flatland), was represented in WorldCat in more than 70 editions, including translations into 6 languages, and led to 4 new works. Each of the four databases searched in this project (WorldCat; Orion for copies in local UCLA libraries; Melvyl for University of California holdings, and Library of Congress holdings) arranges entries differently. For example, the first entry in WorldCat was the one that was the most frequently found in member libraries, the piece of information that end users most likely do not regard as significant in their searching. For example, the edition published by Barnes & Noble in 1963 is held by 863 libraries. The next entry represents the Dover edition published in 1952; 760 libraries own it, etc. Moving through the UCLA’s library catalog to identify local holdings of the Dover edition published in Pasadena by Grant Dahlstrom in 1978. The progenitor was at the bottom of the list, record #9.

It was unclear how these records were arranged in the respective lists. Informal probing of students who searched the Orion library catalog revealed their confusion with regard to sequencing of entries that was presented to them. The response to my inquiry about the Melvyl display convention reads: “In the Melvyl Catalog, records are displayed by default in a ‘last in, first out’ order.” California Digital Library Services (CDLS) go on to explain that this sort of display is “standard for many online catalogs.” They write, “Basically, the most recently loaded records in the database will display first on the screen” (CDLS, personal communication, April 4, 2005). In short displays of up to two screens, this may not be as bothersome for users to scan lists of records briefly as it might be in long search results. Carlyle and Summerlin (2002) write that displays of hundreds or even thousands of records “do little to shed light on the nature and characteristics of the records retrieved.” In contrast, “displays that organize retrieved record sets into intelligible categories may communicate search results more quickly and effectively to users than current catalog displays . . .” (p. 3). For users who have no patience to scroll through multiple screens, the first displayed entry might be the one they would select, ignoring all other editions that are available to them in a given collection.

**Types of Derivative Bibliographic Relationships in Flatland**

The WorldCat database (n = 86) contained the following types of derivative expressions.

- **Amplified editions.** Of 86 bibliographic records of Abbott’s *Flatland*, 79 records were clustered under the same work, and subdivided into the English-language material (n = 70) and translations to six non-English-language materials (n = 9). Of the English-language materials, 26 editions contained introductory texts by various writers (e.g., W. Garnett, B. Hoffmann, I. Asimov, K. Feiden, R. Bradbury, D.V. Davies, A.K. Dewdney, I. Stewart, A. Lightman, and T. Banchoff). In the subset of these editions, 16 are both amplified and numbered. There are 10 amplified editions alone (containing additional introductory texts); 9 of the 10 editions transcribed personal names in a subfield of the statement of responsibility, 245 $c, and once in the notes field (500) with an added entry in the 700 field. The general note reads, “With an introd. by Ray Bradbury.” However, the writer’s name was not given in the statement of responsibility field. In the amplified expressions that are also coded as numbered or consecutive editions (n = 16), personal names were given five times in the edition statement field 250.

- **Numbered or successive editions.** In the subset of 70 English language materials, numbered editions were indicated with wording such as “nth ed. rev.” “1st Shambhala ed.,” “New nth ed.,” and “New and rev. ed.” Numbered editions, as mentioned above, are not mutually exclusive with amplified editions. All numbered editions gave edition information in the edition statement field 250.

- **Translated editions.** Indicative of translated editions (n = 9) is the presence of the fixed field 008/35–37 with a three-letter language code excluding “eng.” Other fields are 041 1_, 240 $l, 245 $c, 250, 260, 5xx, the language note field (546), series statement 490 or 440, and added entry for names of translators, 700 (55%, n = 5). Abbott’s *Flatland* is translated to Italian, French, Greek, Hebrew, Persian, and Russian.

To cluster translated editions as expressions automatically, the only reliable field would be 008/35–37. Physical format in the 006 MARC field informs us if an item is printed language material. The 008 field contains a mandatory three-letter code for language materials. In addition, all records but one coded language information in the 041 field. Six of nine records gave language information in the uniform title field 240. Only three of nine records gave the title in English in the title field 245. For example, the Russian translation transcribes its title *Flatlandija* in the 245 field; the Italian translation uses their own *Flatlandia*, the Hebrew translation uses *Shattuahlandiyah*, and the Persian uses *Pakhshistaan* in the 245 field.

**Other bibliographic relationships.** These three major derivative bibliographic relationships exhibited additional relationships. All editions were illustrated and given in the 245 $c, the 300 $b, and often in the notes field. In addition, equivalence was present in the form of reprints, microforms, and electronic resources. Indication of “form of item” is given in the fixed field 008. Microforms and microfiche are coded...
with a single letter “a” and “b,” respectively. Large print material is indicated with a “d,” regular print reproductions with an “r,” and electronic items with an “s.” Physical description is coded in the physical description field 300; for two microform and four electronic records in the database under study, the extent of carrier is given in fixed field 007 and transcribed in the companion 300 field. Simultaneous derivations are defined as those works “that are published in two editions simultaneously, or nearly simultaneously” (Smiraglia, 1992, p. 28). In Flatland, several such editions were presented in the sample of 86 records. For example, the editions by Little, Brown (Boston) and Blackwell (Oxford) were both published in 1926. William Garnett wrote the introductory note for both editions; however, the editions varied in physical description, and Blackwell’s edition was noted as the “3rd ed. rev.” Mapping of MARC21 linking entry fields 760–78X to FRBR and Tillett’s bibliographic relationships has been discussed in Riva (2004).

Some amplified editions are part of a series (490, 440), such as Princeton Science Library (Princeton University Press, Princeton, NJ), Dover Thrift Editions (Dover Publications, New York, NY), Penguin Classics (Penguin Books Ltd., London, UK), and Project Gutenberg (a producer of free online ebooks, www.gutenberg.org). In nine translations, five are part of a series. These are all considered as hierarchical or whole-part relationships. Shared relationships, defined as works that share an attribute such as subject or title, are also present in three English-language editions and in one translated edition. Information for shared relationships is given in the general notes field and worded, e.g., “Followed by Dionys Burger’s Sphereland.” One record, using full cataloging, describes a shared bibliographic relationship in a subfield 245 $c (“Sphereland: a fantasy about curved spaces and an expanding universe/ by Dionys Burger; translated from the Dutch by Cornelie J. Reinboldt; foreword by Isaac Asimov.” A general note mentions “two separate books, each with forwards by Isaac Asimov. This edition with the texts bound back to back and upside down.” Again, it is noted (in a 500 field) that “Flatland was originally published in 1880. Sphereland was originally published, in Dutch, in 1965.” Personal names for the writer, Burger, and the translator from Dutch, Reinboldt, are given added entries.

New works. Of the new nonbook formats, indicative coding includes letters “m” for motion picture, and “v” for video recording, in the fixed field 008. Fixed field 007 uses “v” for video recording and motion picture, as well as specific characters to code color, format, sound on medium, medium for sound, dimensions, and playback channels. For motion pictures, there is additional coding for production elements, generation, base of film, kind of color stock or print, deterioration stage, and completeness. The AACR2R identifies physical format in the general material designation (GMD; MARC subfield 245 $h); however, this is optional and may be treated differently in different Anglo-American cataloging traditions. The presence of other fields that are indicative of nonbook materials are 260, 300, 5xx, 650 (genre), and personal and corporate name added entries (700, 710).

Discussion

Inconsistent treatment of important contributors to a given work by transcribing them into various fields (245 $c, 700, 500, and 250) makes it difficult to collocate expressions automatically. Furthermore, there are presently no separate tags to transcribe various authoring responsibilities in a 245 $c. A computer algorithm would have to match the presence of truncated text strings for authoring responsibilities. Examples in the set of 79 records under study are “with introd# by,” “with ill# by,” “forward# by,” “with notes by,” and “includes bibl# reference#.” Numbered or consecutive editions collocate with the edition statement field 250. Translated editions to non-English languages use other than “eng” three-letter code in the fixed field 008/35–37. Six of nine records have a 240 field.

The main types of expressions found in the 86 records under study are formalized below as follows:

\[ \text{EXP: } A | B | C | D | E | F | \text{EXP AND EXP} \]

where A is the amplified edition; B is the numbered edition; C is the translated edition; D is the illustrated edition; E is the with reference notes; and F is the simultaneous or near simultaneous edition.

Given ABD, the grammar rule above could be used to parse the expression to inform us if the expression ABD is valid:

\[ A \rightarrow \text{EXP \{} \text{EXP} \} \]

\[ B \rightarrow \text{EXP \{} \text{EXP} \} \]

\[ D \rightarrow \text{EXP \{} \text{EXP} \} \]

According to FRBR, expressions are embodied in manifestations, and these are exemplified in items. Looking at the display of records for multiple manifestations of the same work (from http://www.loc.gov/marc/marc-functional-analysis/functional-analysis.html), it is already possible to display translations of Abbott’s Flatland automatically. Because the FRBR has not been implemented into cataloging rules or extended into the MARC format, collocation of other types of derivative relationships has not been implemented in currently operational OPACs. Expressions and manifestations are typically not distinguished in current cataloging practice. They are all treated as different manifestations of a work.

Concluding Remarks and Further Work

We need to design mechanisms that would guide users in browsing and searching resources that are linked together by explicitly expressed relationships. The metadata schema must be able to link multiple versions of the same resource, to show how these resources are interrelated explicitly, and
how these are related to other similar resources in a digital collection. For example, links could be constructed to show that:

For new works:
- $X$ is based on $Y$
- $X$ is annotated $Y$
- $X$ is related to $Y$
- $X$ is animated interpretation of $Y$
- $X$ is film adaptation of $Y$
- $X$ is computer animation film of $Y$
- $X$ performs $Y$

For expressions:
- $X$ is expression of $Y$
- $X$ is numbered edition of $Y$
- $X$ is new edition of $Y$
- $X$ is revised edition of $Y$
- $X$ is translation of $Y$
- $X$ explains $Y$ (contains preface)
- $X$ contains illustrations for $Y$
- $X$ is abridged edition of $Y$
- $X$ is enlarged edition of $Y$
- $X$ is named edition of $Y$

For manifestations:
- $X$ is release of $Y$
- $X$ is rereleased $Y$
- $X$ is published edition of $Y$
- $X$ varies in date from $Y$
- $X$ varies in pagination from $Y$
- $X$ varies in layout from $Y$
- $X$ varies in typesetting from $Y$
- $X$ is microform of $Y$

For equivalence relationships:
- $X$ is microfiche of $Y$
- $X$ is reprint of $Y$
- $X$ is electronic version of $Y$

For whole–part relationships:
- $X$ is part of $Y$
- $X$ is included in $Y$

For sequential relationships:
- $X$ follows $Y$
- $X$ precedes $Y$

For shared relationships:
- $X$ has same title as $Y$
- $X$ has same author as $Y$
- $X$ has same topic as $Y$

Entries should be arranged in a helpful way to facilitate choice of a desired manifestation of a given work. Whether it concerns assembling all expressions and manifestations of Abbott’s *Flatland*, or XML programming manuals, or Knuth’s *The Art of Computer Programming*, or Shakespeare’s *Romeo and Juliet*, searchers in GDLs ought to see a pathway representing a work along with all associated expressions. Current and future catalogs should build this sort of capability to make participation between systems and end users more collaborative and symmetrical than is currently possible.

Linking of entities based on derivative bibliographic relationships would be particularly useful for technical, environmental, earth sciences, pharmaceutical, and biomedical literature where resources are likely to evolve over a long time period within large projects, interdisciplinary teams, and among distributed laboratories. The reviewed literature revealed no published studies that examined the nature of bibliographic relationships between and among such entities in digital repositories. This author is currently examining the relative importance of weights between parts of entities: text files, source codes, interim solutions, draft case studies, experimental data, illustrations and drawings, and annotated versions. For example, is source code the central focus of a report, or an example that is used to support a thesis of the report?

The importance of understanding the publication life cycle among technical documents, has been well documented (Allen, 1979; Garvey, 1970), and applied in the Engineering Information Sources and Access Project from the perspective of displaying better information resources for engineering students (Ercegovac, 2001). Findings from this study may shed light on organizing retrieval displays by content relationships between and among entities and displaying search results taking into account research from bibliographic families, record clustering, and human information processing.

**Acknowledgments**

This author wishes to thank Dorothy McGarry of UCLA for reading the first draft of this manuscript, for asking questions, and offering useful comments. Thanks must go to Vuk Ercegovac of the University of Wisconsin for useful discussion on issues in data modeling. Finally, the author wishes to thank the anonymous reviewers who provided much appreciated editorial advice.

**References**


