

International Conference on the Principles and Future Development of AACR

Toronto, Canada, October 23-25, 1997

Bibliographic Relationships

by

Sherry L. Vellucci

*Assistant Professor, Division of Library and Information Science
St. John's University, New York*

Bibliographic Relationships

**A Paper Prepared for the
International Conference on the Principles and Future Development of AACR
University of Toronto Faculty of Information Studies
October 23-25, 1997**

**by
Sherry L. Vellucci**

Table of Contents

Introduction.....	1
Background and Context	1
Types of Bibliographic Relationships	4
Relationships Defined by Current Research	5
Relationships Defined by the IFLA Study Group on the Functional Requirements of the Bibliographic Record	9
Generalized Level Relationships.....	11
Lower-level Relationships	12
User Needs: Navigating the Relationship Universe.....	14
Dual Function of Linkages	14
Current Linkage Systems.....	15
Linkage Types and Associated Relationships.....	16
Relationships Important to Catalog Users.....	19
User Studies	20
IFLA Study	22
Relationships Important to Authority Record Users.....	23
Improved Methods for Expressing Bibliographic Relationships in the Online Environment	26
The MARC Environment.....	26
New Catalog Environments.....	28
Local vs. Global Bibliographic Universe	29
General Principles for Bibliographic Relationships in Catalogs	31
Notes	33
Bibliography of Works Cited.....	36

Introduction

Bibliographic relationships exist when bibliographic entities—i.e., any instances of recorded knowledge—are associated with each other in some way. For more than a century, bibliographic relationships were discussed by implication within the context of the collocating function of the catalog. It was not until the late 1970's, however, when the move toward online catalogs sparked hopes for a more sophisticated catalog structure, that bibliographic relationships became an independent topic of discussion.¹ Theoretical and empirical examination of the concept began in earnest in the 1980's and continues to the present day.

For the purpose of defining the role of bibliographic relationships within the principles and future development of the *Anglo-American Cataloging Rules* (AACR), it is important to understand the types of relationships that exist in the bibliographic universe as they have been identified to date, and the means currently used for identifying and linking bibliographic records for related bibliographic entities. It is also crucial to have a contextual understanding of this topic as it relates to the functions of the catalog; the functions of the bibliographic record; the concept of the work; the concepts of main entry, added entries, and references; bibliographic and authority record structure; and catalog design. As several of these topics will be presented as independent position papers at this conference, there is no need to explore them all in depth. But it is important to comprehend that bibliographic relationships do not exist as an isolated concept. They are enmeshed in almost every aspect of the cataloging process, are alluded to throughout the AACR, and are an important factor in the structure and design of the catalog.

Bibliographic relationships should also be distinguished from other relationships that exist in the bibliographic universe, e.g., the relationship of one name to another, or the relationship of one subject to another. While both names and subjects may be related to a bibliographic entity, and both name entities and subject entities may be used to help identify bibliographic relationships and link related records, it is primarily those relationships between two bibliographic entities that are included in the study of bibliographic relationships.

Background and Context

In her ground-breaking study of bibliographic relationships, Tillett noted that in the library catalog, established bibliographic relationships are pathways that provide a systematic display of related materials. These pathways direct us to associated materials, and that direction “helps to fulfill the stated objectives of the library catalog” (Tillett, 1987, 2). Because identification of bibliographic relationships is integral to the fulfillment of the objectives of the catalog, a brief discussion of these objectives as they appertain to bibliographic relationships is appropriate.

One of the first detailed discussions regarding the functions of a library catalog was occasioned by Panizzi's need to defend the systematic cataloging rules that he developed to create a catalog of the holdings of the library of the British Museum. In the hearings before a Royal Commission, Panizzi stated that “A reader may know the *work* he requires; he cannot be expected to know all the peculiarities of different *editions*...” (Panizzi, 1850, 695; emphasis in the original). Panizzi strongly believed that it was the job of the catalog to collocate these related editions of the work, so that the user had all information to hand when making a selection. In 1848 Panizzi posited five basic characteristics for a library catalog: it should provide enough detail that users could identify a specific book, it should have only one entry for each bibliographic item, it should use a normalized form of an author's name for that entry, it should arrange all the editions and translations of a work together, and it should provide cross-references to lead users to the appropriate forms of names and titles (Panizzi, 1985). It was Panizzi's fourth and fifth

characteristics—i.e., the collocating function for related works, and references to appropriate titles—that linked bibliographic relationships with the functions of the library catalog.

Thirty years later, Cutter clearly stated the objects of the catalog in his landmark work *Rules for a Printed Dictionary Catalog*. These objects, which are ubiquitously quoted today, read:

- 1). To enable a person to find a book of which either
 - a) the author is known
 - b) the title is known
 - c) the subject is known;
- 2). To show what the library has
 - d) by a given author
 - e) on a given subject
 - f) in a given kind of literature;
- 3). To assist in the choice of a book
 - g) as to its edition (bibliographically)
 - h) as to its character (literary or topical) (Cutter, 1904, 12).

The objects of a catalog, as detailed by Cutter, essentially served the same identifying and collocating functions that were delineated by Panizzi, and added the evaluating function that was implicit in much of Panizzi's writing. Cutter's second object, collocation, depends upon gathering together material that is related by virtue of author, subject, or form. He does not, however, include gathering all editions of a work under this rubric. Nevertheless, this first set of explicit stated objects did signal that identifying bibliographic relationships and linking records for related materials were intrinsic functions of a catalog, and marked the beginning of the "second objective" as the focal point for those concerned with collocation and bibliographic relationships.

Almost half a century later, the Library of Congress (LC) asked Lubetzky to examine and define the underlying principles of cataloging that would form the basis of a new and simplified code. As a starting point Lubetzky examined the functions of the catalog, for he believed that the defined functions of the catalog would determine the direction and purpose of any new code. In his draft, Lubetzky cited the following two purposes that the catalog should serve: "First, to facilitate the location of a particular publication, i.e., of a particular edition of a work, which is in the library. Second, to relate and display together the editions which a library has of a given work and the works which it has of a given author" (Lubetzky, 1960, ix). Lubetzky's functions did not include the treatment of subjects or form. While his functions were more sophisticated than Cutter's in that they introduced the concepts of the item and the work as separate entities and called for collocating various manifestations of the work, they did reiterate Cutter's objects of identifying and collocating. The evaluating function was implied but not explicitly stated. Although a final code was never published, Lubetzky's work served as a basis for the *Statement of Principles* adopted at the International Conference of Cataloging Principles (ICCP), Paris, 1961 (ICCP, 1963, 91-96), known thereafter as the "Paris Principles."

The developers of the "Paris Principles" reaffirmed that the catalog served a dual purpose: it identified and located specific items (the identifying, or finding function) and enabled related items to be identified and located (the collocating function). The following objectives were included in the report of the "Paris Principles":

The catalogue should be an efficient instrument for ascertaining

1. whether the library contains a particular book...and
- 2. which works by a particular author and which editions of a particular work are in the library (ICCP, 1963, 91).**

Again it is the “second objective” that focuses on the collocation of related items and works. Since the 1961 Conference, each successive publication of the *Anglo-American Cataloging Rules* has acknowledged Lubetzky’s contribution to the formulation of cataloging principles, but the ICCP objects of the catalog have never been explicitly incorporated in the code.

More recently, Svenonius noted that the three decades since Lubetzky reformulated the catalog’s objectives have witnessed “more dramatic changes to the cataloging infrastructure than the preceding seven. Not only have catalogs been automated but also an unprecedented amount of cooperative cataloging has led to the emergence of international standards, global catalogs, and linked systems” (Svenonius, 1989, 2). Expanding on Svenonius’ reference to global systems, O’Neill and Visine-Goetz stated that “the information that was sufficient to identify particular items and bibliographic relationships within a local collection is inadequate in this new [global] environment” (O’Neill, 1989, 172). They questioned whether the objectives of the catalog as stated in the “Paris Principles” are still valid for future library catalogs within the broader context of a global environment.

Wilson responded to this question not by proposing new objectives, but by suggesting a different *priority* for the established objectives (Wilson, 1989a). He maintained that through technological advances, we are moving in a direction where the storage of documents will be separated from their medium of display, and libraries will provide access to what he termed *virtual copies* that are not actually present in the local collection, but can be downloaded from some distant database (Wilson, 1989a, 6). In view of this global access, Wilson believed that a reordering of the relative priorities of the objectives of the catalog was in order. Since the time of Cutter, priority had been given to the first objective, i.e., the identifying function, which provides access to the specific physical item in the library. Wilson deemed this priority more valid for an emphasis on a local collection, but believed that the *work*, rather than the physical *item*, was of primary bibliographic interest in the global online environment where the physical item had little meaning. Thus, he postulated that the second objective, i.e., providing access to the work through collocation, should receive first priority. It is the identification of bibliographic relationships and linking of related records that furthers this second objective of collocation.

The next step is to determine whether the functions of the catalog will hold as appropriate for the restructured bibliographic systems of the future. Within the context of restructuring the catalog, Green’s seminar class conducted a requirements analysis to determine the functions of a bibliographic database to be designed using an entity-relationship model. The resulting statement provided the following delineation of the purposes of a bibliographic database:

Identification

Verify that a bibliographic unit exists.

Verify information about a bibliographic unit so as to make accurate and distinguishing reference to it.

Ascertain whether a bibliographic description matches a bibliographic unit.

Ascertain whether two bibliographic units are the same

a) bibliographic entity (“book”)

b) intellectual entity (“work”)

Selection

Determine the (absolute) appropriateness to a user’s situation of

a) a bibliographic entity

b) an intellectual entity

Determine the (relative) appropriateness to a user’s situation of two

a) bibliographic entities

b) intellectual entities

Availability

Determine the physical location of a bibliographic unit.

- Determine (restrictions on) availability (e.g., hours facility is open, document delivery delay, required user authorization) of a bibliographic unit for
- a) local use.
 - b) loan (including interlibrary loan).
 - c) purchase.

Relationships (Based on Tillett, 1991)

- Determine the relationship between a work and any of its bibliographic manifestations.. (“equivalence”).
- Determine the relationship between a work and any works produced by modifying it... (“derivative”).
- Determine the relationship between a work and a work that describes it ... (“descriptive”).
- Determine the relationship between an intellectual or bibliographic entity and its components... (“whole-part”).
- Determine the relationship between a bibliographic unit and another bibliographic unit that accompanies and either complements or augments it... (“accompanying”).
- Determine the relationship between a work and a predecessor work... (“sequential”).
- Collocate works or bibliographic units that have some characteristic... in common... (“shared characteristic”) (Green, 1996, 209).

Examination of this statement shows that the overall functions remain much the same as the traditional catalog functions, but the details have been altered to incorporate a clear distinction between the item and the work, to expand the concept of availability, and to delineate specifically the concept of expressing relationships. This new statement enables the functional requirements of a catalog to be discussed within the context of a new database structure and the broader global environment.

The adequacy of the traditional catalog functions still remains to be decided by the cataloging community. It appears, however, that it is not the functions themselves that require revision—although rewording and redefining may be useful—as much as the objects of these functions: the physical item versus the abstract work, and the local library collection versus the global electronic library environment. Should the catalog continue to represent only the physical holdings of a given library’s collection? In fact, with the addition of bibliographic records for remotely accessed electronic resources and consortia holdings, catalogs already go beyond the limited scope of local holdings. The real questions to be answered are “what are the parameters of the catalog?” and “what level or levels of description and access, vis-a-vis the item or work, should the catalog include?” In order to accommodate bibliographic relationships within a global environment, the parameters of the library catalog should be expanded to include records for those entities to which the library wishes to provide access; and, if new database structures are to be considered seriously, the catalog must provide description and access at the abstract level of the work, as well as the lower level of the physical item.

Types of Bibliographic Relationships

While the need to clarify relationships among bibliographic entities in library catalogs has been acknowledged for many years, the formal study of bibliographic relationships is relatively new. One of the first attempts to define and categorize types of bibliographic relationships can be found in the *UNIMARC Format* (1980) published by the International Federation of Library Associations and Institutions (IFLA). The format identified three main categories of relationships:

Vertical—the hierarchical relationships of the whole to its parts and the parts to a whole ...;
Horizontal—the relationship between versions of an item in different languages, formats, media, etc. ...;
Chronological—the relationship in time between issues of an item ... (UNIMARC, 1980, 58-59).

The IFLA committee that developed the *UNIMARC Format* approached the identification of bibliographic relationships at a highly theoretical level. While these three broad categories are comprehensive in that all relationships can be fitted into them, they do not contain the degree of detail necessary to describe, distinguish, and classify all types of more complex bibliographic relationships.

Goossens and Mazur-Rzesos offered a second attempt at the analysis of relationships, but limited their examination to whole-part relationships, identified as vertical by the UNIMARC document (Goossens, 1982). In this study Goossens and Mazur-Rzesos created a three-tier schematic representation to express both simple and complex hierarchical relationships as tree structures. Their objective was to create a theoretical model for expressing such relationships in both manual and automated systems. The limitations inherent in confining the model to one type of relationship, however, do not allow it to be used for non-hierarchical relationships such as the horizontal relationships identified by the *UNIMARC* publication. Thus, this type of analysis cannot be used to identify and classify all types of bibliographic relationships.

Relationships Defined by Current Research

Recent research into bibliographic relationships has had a more practical focus, identifying specific relationships in detail, investigating the frequency with which bibliographic relationships occur, and examining the ways in which these relationships are explicated in the catalog. While each study made its own unique contribution to the knowledge base, it was Tillett's vanguard research of bibliographic relationships that succeeded in presenting the first detailed analysis of relationship types (Tillett, 1987). Her study began with the creation of a taxonomy of bibliographic relationships that she discovered by examining 24 different cataloging codes from Panizzi's 1841 *Rules* to the 1978 second edition of *AACR*. Tillett's definitions for the seven categories of relationships identified in her taxonomy are presented in the following list, along with a summary of her examples for each category:

Equivalence Relationships, which hold between exact copies of the same manifestation of a work, or between an original item and reproduction of it, as long as the intellectual content and authorship are preserved. Included here are copies, issues, facsimiles, photocopies, microforms, and other similar reproductions;

Derivative Relationships, called horizontal relationships in UNIMARC, which hold between a bibliographic item and a modification based on that same item, including variations, versions, editions, revisions, translations, adaptations, paraphrases, etc.;

Descriptive Relationships, which hold between a bibliographic item or work and a description, criticism, evaluation, or review of that work, including annotated editions, casebooks, commentaries, critiques, etc.;

Whole-Part Relationships, called vertical relationships in UNIMARC or hierarchical relationships by Goossens, which hold between a component part of a bibliographic item or work and its whole, including a selection from an anthology, collection, or series;

Accompanying Relationships, which hold between a bibliographic item and the bibliographic item it accompanies, such that the two items augment each other equally or one item augments the other principle or predominant item, including supplements, concordances, indexes, catalogs, etc.;

Sequential Relationships, called chronological relationships in UNIMARC, which hold between bibliographic items that continue or precede one another, including successive titles in a serial, sequels of a monograph, parts of a series, etc.;

Shared Characteristic Relationships, which hold between a bibliographic item and other bibliographic items that are not otherwise related but coincidentally has a common author, title, subject or other characteristic used as an access point (Tillett, 1987, 24-25).

In addition to bibliographic relationships, Tillett examined the cataloging codes to identify the various types of linking devices used to establish each type of relationship on the bibliographic record. These will be discussed later in the section “Current Linkage Systems.”

The second part of Tillett's work included an empirical study designed to examine the extent of bibliographic relationships as reflected in their frequencies of occurrence in MARC records entered in the Library of Congress machine-readable database between 1968 and July 1986. One of the important findings of this portion of the study indicated that bibliographic relationships are widespread throughout the bibliographic universe, for Tillett found that almost 75 percent of the records in the database contained some type of relationship information (Tillett, 1987, 190). The study included records for materials in a variety of MARC formats, including books, serials, maps, visual materials, and music. Although the overall proportion of Tillett's sample exhibiting each type of relationship is not known, she does provide figures that represent the proportional breakdown of all relationships found. Thus, 62.32 percent of the relationships were whole-part, 16.14 percent were sequential, 14.29 percent were derivative, 3.91 percent were accompanying, and 3.34 percent were equivalence. When distribution by format was examined, it was found that the accompanying relationship predominated among visual materials and music materials, with high proportions of sequential and equivalence relationships found among serials, and whole-part and derivative relationships predominating among maps. This particular discovery was significant for it suggested that different formats, and in some cases different disciplines, will have a propensity for certain relationship types, and, therefore, care must be taken not to make general statements about relationships that are based on observations of largely monographic collections.

Tillett's study had two major strengths. First, her taxonomy was derived from examination of the cataloging rules, thus identifying those relationships that fell within the scope of the codes' underlying principles and merited treatment in the rules. Second, her study extended to constructing a profile of the means provided by the cataloging codes to identify and link the specific relationships in the catalog. With this code-based taxonomy in place, there was still a need for empirical research to verify these relationships in the bibliographic universe, and to document further the extent of these relationships in the real library environment.² Tillett's landmark study became the starting point for two further investigations that focused on the bibliographic universe represented in library catalogs, both of which were narrower in scope.

The study conducted by Smiraglia, examined bibliographic families found in the academic library catalog of Georgetown University. The purpose of the study was to discover the characteristics and extent of derivative relationships among works (rather than items) represented in an online catalog (Smiraglia, 1992). In an article describing his dissertation research, Smiraglia defined derivative relationships as those “bibliographic relationships that exist between any new conception of a work and its original source, or its successor, or both” (Smiraglia, 1994, 172).

After examination of the AACR, he defined the following seven categories of derivative relationships:

Simultaneous derivations. Works that are published in two editions simultaneously, or nearly simultaneously... Often such simultaneous derivations will exhibit slightly different inherent bibliographic characteristics;

Successive derivations. Works that are revised one or more times, and issued with statements such as “second, [third, etc.] edition,” “new, revised edition,” works that are issued successively with new authors, as well as works that are issued successively without statements identifying derivation;

Translations, including those that also include the original text;

Amplifications, including only illustrated texts, musical settings, and criticisms, concordances and commentaries that include the original text;

Extractions, including abridgements, condensations, and excerpts;

Adaptations, including simplifications, screenplays, librettos, arrangements of musical works, and other modifications;

Performances, including sound or visual (i.e., film or video) recordings (Smiraglia, 1994, 177).

He later added an eighth category—*predecessor derivations*, that included the work from which the progenitor clearly derived. Smiraglia went on to suggest that it was this network of related works that constituted a bibliographic family, which he defined as the “accumulation of works that deliberately share ideational and semantic content, and that are derived from a progenitor work” (Smiraglia, 1994, 172).

Smiraglia’s study contained several important findings for the study of bibliographic relationships. First, 49.9 percent of the works in his sample exhibited a derivative relationship, a proportion that is more than twice as large as the percentage discovered by Tillett for this relationship category (Smiraglia, 1992, 60). Smiraglia directly attributed this discrepancy to methodological differences between the two studies (Smiraglia, 1992, 60). Specifically, he examined the documents themselves, rather than relying exclusively on information found in the bibliographic records, and he examined the bibliographic universe beyond the catalog of one library’s holdings. Second, data from Smiraglia’s study suggested that derivative relationships were most strongly associated with the characteristic of age of the progenitor work—i.e., bibliographic families with older progenitors tended to be larger. This has implications for catalogs of libraries with large humanities collections where older literature remains relevant for longer periods of time than scientific literature. Third, between 40 percent and 63 percent of the derivative relationships found to exist in the study did not have any explicit linkages in the Georgetown catalog. In other words, the existence of the known relationship was not indicated on the record, nor could it be inferred by the user from collocated records (Smiraglia, 1994, 175). Smiraglia attributed this, in part, to the progenitor work not having information relating to derivative works that succeeded it. In addition, he noted that unless a given library owned all manifestations of a work, it was impossible to identify all derivative relationships from that one catalog. Smiraglia observed that “when the catalog of any institution contains only a portion of a bibliographic family, implicit linkage is even more likely to fail to make the relationship apparent” (Smiraglia, 1992, 61). The results of this study again lead back to questioning the parameters of

the catalog and levels of access required to provide the user with explicit linkages among related bibliographic works, rather than relegating this function to the users deductive powers.

A third study conducted by Vellucci, focused on the bibliographic relationships that exist among musical bibliographic entities represented in library catalogs (Vellucci, 1995). The basis for the study was a sample of music scores drawn from the catalog of the Sibley Music Library, Eastman School of Music. This research included three major goals: to identify general characteristics of music scores found in a library collection; to describe and classify the bibliographic relationships that exist among musical bibliographic entities and other bibliographic entities found in the bibliographic universe of library catalogs; and, to identify and categorize the linkages used to identify and link related musical bibliographic entities in the library catalog. In order to provide contextual rather than theoretical information, the research design used existing musical entities as the basis for determining the bibliographic relationships.

Vellucci identified the same relationship categories for music as those found by Tillett, but the subgroups within each category varied because of the nature of musical entities. The following six categories of bibliographic relationships were defined based upon their application for musical entities:

Equivalence Relationships, that exist between exact copies of the same manifestation of a work, or between an original item and a reproduction of it, when the reproduction is intended to function as a substitute for the original. Subgroups include exact copies; microform copies; manuscript reproductions; issues, reissues, and impressions; and photocopies;

Derivative Relationships, that exist between any new conception of a work and its original source (the progenitor), or a successor, or both. Subgroups include performances (audio and visual); derivative editions; amplifications; adaptations; arrangements; forms of musical presentation; translations; and notational transcripts;

Descriptive Relationships, that exist between a bibliographic entity or work and a description, criticism, or evaluation of that work, including annotated editions, commentaries, analyses, etc.. Subgroups include description included with printed musical work; separate text; audio description; visual description; filmstrips, kits, etc.; and performance programs;

Whole-Part Relationships, that exist between a segment or component part of a bibliographic entity or work and its whole. Subgroups include physical inclusive; physical extractive; abstract inclusive; and abstract extractive;

Accompanying Relationships, that exist between the primary musical item or work and the complimentary material that occurs with it, such that the two augment each other equally, or the secondary material augments the primary musical item or work. Subgroups include supplemental and inclusive; and,

Sequential Relationships, that exist between bibliographic items or works that precede or continue one another, embodying the aspects of chronology and sequence, but are not derivative. Subgroups include series, sequel, and serial (Vellucci, 1995, 310-311).

Among the more important findings of her study was the high degree of relatedness found among musical entities. Vellucci discovered that 97 percent of the scores in the sample exhibited at least one relationship, a considerably higher figure than that discovered by Tillett. This was probably due to several reasons including the methodological differences identified by Smiraglia that also

held for Vellucci's study, and factors stemming from the inherent nature of music that requires performance for its aural realization. Included among the musical factors are the concept of the *performable unit*, which encourages sections of a musical work to be published separately; the need for performing editions and performance parts; the preponderance of western art music standard repertoire and its continued demand regardless of age; and specific genres of music that tend to be produced and published in groups, anthologies, or collections. In addition, the international scope of music, and the music publishing industry with its focus on specific sales markets, contribute to the extent of relatedness. Not surprisingly, therefore, the whole-part, derivative, and accompanying relationship categories each had a high frequency of occurrence, exhibited by 86 percent, 85 percent, and 71 percent of the sample, respectively. The remaining three categories of relationships examined in this study—sequential, equivalence and descriptive—were discovered for 31 percent, 29 percent, and 22 percent of the sample, respectively. Vellucci also found that 33 percent of the sample exhibited some type of relationship that was not identified on the bibliographic record. In addition, the age of the composition was found to be the only characteristic that had a statistically significant association with every type of relationship (although other characteristics showed significant associations with specific relationships). This finding supported that of Smiraglia, i.e., the older the literature, the greater the chance of bibliographic relationships existing.

Several observations can be generalized from these three studies. The seven types of relationships identified by Tillett's conceptual analysis of cataloging codes—i.e., equivalence, whole-part, derivative, accompanying, sequential, descriptive, and shared characteristics—appear to be valid in the bibliographic universe; bibliographic relationships are widespread among items and works found in the bibliographic universe; and, the most reliable predictor of relationships in general appears to be the age of the work, although other characteristics may be associated with specific relationships. It also appears that certain categories of entities—e.g., music scores, serials, and non-print materials—tend to exhibit a higher than expected frequency of relationships. In addition, it seems that not all existing relationships are identified and linked in the catalog by our current methods of creating, storing, and displaying bibliographic records. This last aspect of bibliographic relationships will be discussed below in the section “User Needs: Navigating the Relationship Universe.”

Relationships Defined by the IFLA Study Group on the Functional Requirements of the Bibliographic Record

The IFLA Study Group on the Functional Requirements of the Bibliographic Record adopted a different approach to the examination of bibliographic relationships than the empirical research just discussed (IFLA, 1996). First, it used the functions of the bibliographic record as the starting point for analysis, rather than the functions of the catalog. Second, the IFLA study used an entity-relationship model to analyze the bibliographic record, thus structuring the study to emulate recent research into relational database and object-oriented database modeling for online catalogs.³ Previous bibliographic relationship research provided a descriptive analysis of relationships, but did not attempt a structural model. Third, the study was broader in scope than the previous studies, in that it first included a detailed analysis of three entity groups described in the bibliographic record. These included entities that represent the products of intellectual or artistic endeavor (*works, expressions, manifestations, and items*), entities responsible for the intellectual or artistic content of such products (*persons and corporate bodies*), and entities that form the subject of intellectual or artistic endeavor (*concepts, objects, events, and places*) (IFLA, 1996, 9). The previous studies focused primarily on the bibliographic relationships exhibited by the entities defined as items and works, with the responsible entities included only in the examination of some linkages, and very limited inclusion of subject entities in the discussion of descriptive relationships. Finally, the IFLA Study Group attempted to place a relative value on data in the bibliographic record, basing its weighting on a conceptual analysis of user needs as perceived by the Study Group. Although the previous studies attempted to identify through

empirical research, entity characteristics that could be used to predict the occurrence of specific relationships, no attempts were made to place value on bibliographic data. The user needs for bibliographic data identified by the IFLA Study Group were, in fact, a rewording of the four primary functions of the catalog that have been discussed for over a century—finding, identifying, selecting, and obtaining. This may dispel the criticism that the study did not take the functions of the catalog into account, but will support the criticism that the study did not base user needs—and therefore the relative value of bibliographic data—on information verified by empirical research.

Before discussing the IFLA report's treatment of bibliographic relationships, a few words must be said about the definitions and the terms used to represent the products of intellectual or artistic creation—i.e., *work*, *expression*, *manifestation*, and *item*—for the relationship modeling is based on these entities. Since the document was disseminated for worldwide review, there has been considerable criticism, a portion of which centered on these entity definitions and their application. The problem is perhaps more semantic than substantive, and while the intention was clarity, the result was confusion. At times the application of the definitions does not always appear to be consistent throughout the document (which may be due in part to a lack of clarity and inappropriate examples). More importantly, these definitions confuse because they have modified certain concepts of long standing, and have applied to different entities previously used entity definition terms. For example, the *work*, as referred to in most previous studies of bibliographic relationships, has been defined as the intellectual or artistic content of a bibliographic entity; the *work* consisted of two properties: the ideational content and the semantic content, and existed in the abstract. The IFLA study defines work as “a distinct intellectual or artistic creation” (IFLA, 1996, 11), but does not define the properties of a work. The definition implies, however, that the *work* contains only ideational content, for the property of semantic content appears to be the defining factor for the new second level entity, *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” (IFLA, 1996, 12). The report goes on to state clearly that the *work* is an abstract entity; however, *expression* is not explicitly defined as an abstract entity. Nevertheless, it is stated that the *expression* entity excludes aspects of physical form, thus implying that it is an abstract concept. To obfuscate further the issue, some examples given of the *expression* entity are physical objects, e.g., “the specific score used for the performance of a musical composition” (IFLA, 1996, 13), and the term “object” is actually used in the definition. It is not surprising that there has been a bewildered reaction from the cataloging community.

The confusion is exacerbated because the cataloging community also has been unclear in its definitions and terminological usage. For many years, the *item* has been viewed as a physical manifestation of the work, with the terms *item* and *manifestation* frequently used interchangeably.⁴ The individual exemplar of the physical item has been referred to as a *copy*. The IFLA Study Group has defined the *manifestation* entity as the physical embodiment of an *expression* of a *work* (IFLA, 1996, 14)—what the cataloging community understands alternately as the *item* or the *manifestation*—and has defined the *item* as a single exemplar of a *manifestation* (IFLA, 1996, 15)—what the cataloging community understands as the *copy*. To add to the confusion, other researchers have described this hierarchical entity grouping by various other clusters of terms, including *work*, *text*, *edition*, *printing*, and *book* (O'Neill and Vizine-Goetz, 1989, 172), *impressions*, *editions*, *texts*, *works*, and *superworks* (Svenonius, 1992, 5-6), *uniform work*, *work*, *bibliographic set*, *bibliographic copy*, and a super-entity called *bibliographic unit* (Green, 1996, 210-11), and *text* or *abstract work*, *publication*, and *copy* (Heaney, 1995, 152). Additionally, Yee subdivided the *manifestation* entity into *manifestation*, *title manifestation*, and *near equivalent* (Yee, 1994a, 246). Thus, with the variation in hierarchical cluster terms and definitions that already exists within the cataloging community, the IFLA study has further compounded the semantic problems with its new definitions, murky explanations, and inadequate examples. Since the cataloging code with its glossary is most often referred to as the arbiter in defining cataloging terms, it would benefit the cataloging community to come to consensus on the

concepts and terminology that are so integral to the future development of the descriptive cataloging code. Once agreed to, the definitions could then be incorporated into the future code to provide a common language. In this respect, the IFLA study has furthered the process by providing a detailed document to initiate discussion.

Generalized Level Relationships

The IFLA study defined the relationships used in the model at two different levels, those operating at a generalized level, and those operating at a lower level between specific instances of entities (IFLA, 1996, 7-8). The relationships defined at the generalized level are broader in concept, and incorporate bibliographic relationships as well as other types of relationships that represent logical operations or functions in the entity-relationship model. These generalized level relationships are considered the logical structural framework upon which the specific level relationships are based.

Modeled on the three entity groups, the generalized, or high level accommodates three relationship clusters: relationships between *work*, *expression*, *manifestation*, and *item* (i.e., bibliographic entities); relationships to *persons* and *corporate bodies* (i.e., responsible parties); and *subject* relationships. The relationships are binary—in that they only operate between two entities, and they are bidirectional, in that the relationship, when expressed as its own reciprocal, operates in two directions. The first cluster—relationships between *work*, *expression*, *manifestation*, and *item*—links the hierarchical levels between these entities, each of which incorporates the attributes of its superordinate entity. The following statements illustrate this relationship cluster, with the relationship statements enclosed in angled brackets:

A work <is realized through> an *expression*; an *expression* <is a realization of> a *work*.

An *expression* <is embodied in> a *manifestation*; a *manifestation* <is the embodiment of> an *expression*.

A *manifestation* <is exemplified by> an *item*; an *item* <is an exemplar of> a *manifestation*.

This first cluster can be referred to as the bibliographic entity cluster, for lack of a better collective term. The seven categories of bibliographic relationships identified by previous research operate among this group of bibliographic entities, and may exist at both this generalized level or the specific instance level, depending upon the specific relationship type. Due to the lack of relationship statements in most current catalogs, this relationship cluster employs both descriptive elements and access points on the bibliographic record to identify the relationship at each level of the entity hierarchy.

The second relationship cluster—relationships to *persons* and *corporate bodies*—links the entities in the first group (i.e., bibliographic entities) with the entities in the second group (i.e., responsible parties), using two operational relationships. The following statements exemplify this relationship cluster:

A work <is created by> a *person* or *corporate body*; a *person* or *corporate body* <has created> a *work*.

An *expression* <is realized by> a *person* or *corporate body*; a *person* or *corporate body* <has realized> an *expression*.

Due to the cumulative nature of data in the bibliographic record as entities lower in the entity hierarchy are described, these responsibility relationships for the *work* and the *expression* may be incorporated into the records for *manifestations* and *items*. Other aspects of responsibility also could be included at these two lower levels, if the “responsibility” relationship were redefined to include activities other than creation of the content, to which it is currently limited by definition (IFLA, 1966, 16-17). For example, if the definition were expanded as the Study Group implied it could be under the relationship definition (IFLA, 1996, 47), the following operational relationships may apply:

A manifestation <is published by> *a person or corporate body*; *a person or corporate body* <has published> *a manifestation*.

An item <is owned by> *a person or corporate body*; *a person or corporate body* <owns> *an item*.

The second relationship group is primarily concerned with access points for the bibliographic record, although at the expanded *manifestation* and *item* level, this information is currently contained in the bibliographic description. Additionally, this relationship cluster could be expanded easily beyond bibliographic data, to include, for example, circulation data (<is borrowed by>) and acquisitions data (<is bought from>) in an integrated database model.

The third relationship cluster—subject relationships—links the *work* entity with all the entities included in all the entity groups by the operational relationship <has as subject>. This indicates that any entity may be the subject of a work, including those entities defined as subject—i. e., *concepts*, *objects*, *events*, and *places*—as well as *persons* or *corporate bodies*, and *works*, *expressions*, *manifestations*, and *items*. While the relationship is accommodated in the model because the study examined all aspects of the bibliographic record, subject relationships are generally not part of the descriptive cataloging process, with the exception of descriptive relationships and certain instances of derivative relationships such as when subjects are used as linkages to bibliographic records for authors of texts set to music.

Lower-level Relationships

The second level of relationships operates between specific instances of entities. While these relationships may occur between instances of the same entity type or different entity types, the goal of the study was “to show how the relationships operate in the context of the four primary entities in the model (i.e., *work*, *expression*, *manifestation*, and *item*)” (IFLA, 1996, 48), and therefore the examples were confined to this primary entity group. Again, the relationships are binary, in that they only operate between two entities, and they are bidirectional, in that the relationship, when expressed as its own reciprocal, operates in two directions.

In defining the relationship types, the study deconstructed the seven higher-level relationship categories established by empirical research —i.e., whole-part, derivative, equivalence, accompanying, sequential, descriptive, and shared characteristics—and used instead the more specific subclasses within the relationship categories (IFLA, 1996, 48). For example, the study includes successor, supplement, and adaptation among the relationship types. In fact, successor is a subclass of the higher-level sequential relationship category, supplement belongs to the higher-level accompanying relationship category, and adaptation is a subclass of the higher-level derivative relationship category.

This deconstruction approach has both strengths and weaknesses. Deconstruction is necessary to categorize accurately the nature of the subclasses of relationships, for it is primarily at the subclass level that relationships operate between instances of entities. Each subclass will exhibit its

own unique relationship characteristics, and each will vary as to its entity-relationship statement. For example, <is adaptation of>, <is translation of>, and <is aural performance of> are all types of derivative relationship statements, yet the higher-level relationship statement <is derivation of> is not adequate to explain the exact nature of the relationship. The problem with the deconstruction approach is that the subclasses within each higher-level relationship category are often numerous, and tend to vary depending on the nature of the discipline.⁵ This, no doubt, is one reason why the study does not claim to be exhaustive. Further analysis must be done at this level, both in general terms and within specific disciplines, before a comprehensive entity-relationship model is completed. This deconstruction approach, however, is a great step forward in converting the theoretical higher-level relationship research into a more pragmatic structure useful for database design.

Another useful aspect of the relationship definition model developed by the study is the classification of relationship types by their degree of autonomy. Two categories of relationship autonomy were defined (IFLA, 1996, 50): referential, where one entity is so closely connected to another entity that the relationship has little value outside the context of the other entity; and autonomous, where one entity does not require reference to another entity in order to be useful or understood. While certain relationship types (i.e., successor, supplement, and complement) have greater potential to be included in the referential category than other types, the final decision on autonomy will depend on the characteristics of the specific entity. In addition, it appears that the degree of autonomy applies only among entities at the upper level of the hierarchy—i.e., *work* and *expression*—for autonomy is not discussed in relation to *manifestations* or *items*. The study used this autonomy characteristic in determining the relative importance of identifying a given relationship in the catalog, with greater weight given to those relationships that were dependent, i.e., referential, since the meaningful use of one entity will be dependent on the content of the other entity in the relationship.

The treatment of relationships in the IFLA study is somewhat uneven, for although the report claims to use deconstruction analysis in lieu of describing specific higher-level relationships, one higher-level relationship type—the whole-part relationship—is singled out and discussed separately for each entity type. In reviewing the whole-part relationship, the study identified one important characteristic associated with this relationship type—the degree of dependency of the part to the whole (IFLA, 1996, 52). A part can be either an *independent* component of the whole, in which case it does not depend to any significant extent on the context provided by the larger part for its meaning, or it can be a *dependent* component, that is intended to be used in the context of the larger work and thus depends on the context provided by the larger work for its meaning. The degree of dependency characteristic does not appear to operate at the lower end of the entity hierarchy among *manifestations* and *items*. This may be because the relationship type at these levels is mostly limited to reproductions, which by their nature are equivalent, and thus preclude a dependent relationship. The dependent whole-part relationship is further divided into two subcategories: discrete components whose content exists as an identifiable segment with distinct boundaries, such as chapters, sections, etc.; and integral components without distinctly identified parameters, whose content is interwoven throughout the whole work, such as illustrations (IFLA, 1996, 53).

While in the past it was not usually necessary to identify or create access to dependent parts of a work, this should be re-examined within the context of catalogs that provide direct links to electronic texts, and electronic documents that provide direct access to parts of the work. It may be that access to the whole work is all that is necessary to create the initial linkage, and once linked, the text itself will identify the parts for the user; however, the cataloging of electronic texts should be studied empirically to document the most efficient and effective use of bibliographic records in this environment.

The IFLA study discusses indirectly most of the remaining higher-level relationship types—i.e., derivative, accompanying, descriptive, sequential, and equivalence—through certain of their subclasses; the higher relationships are not mentioned specifically by name. The separate treatment for whole-part relationships may be justified, however, for unlike other relationship types that have several specific subclasses, each of which has its own separate characteristics, the whole-part relationship contains only whole entities or parts of entities, and thus is best discussed at the higher level. As in other bibliographic relationship research, the shared characteristic relationship category is not addressed in the IFLA discussion, for a shared characteristic is ubiquitous and may be found among any entity attributes in all three relationship entity groupings.

In summary, the IFLA study of relationships has provided a model for the structural analysis of data, in order to determine which data elements are necessary to include in bibliographic records that will meet user needs. While problems with the definition of specific entities, attributes, and relationships have created confusion that will engender further discussion, the relationship model succeeds in presenting a cogent structure for converting the theoretical research in the area of bibliographic relationships into a more pragmatic structure useful to various database designs. In order to be useful in the development of a new cataloging code, however, the means used to identify bibliographic relationships in the catalog and the methods prescribed by the rules to link records for related entities must be examined.

User Needs: Navigating the Relationship Universe

From the time of Panizzi onward, syndetic structure became a standard feature of most library catalogs, and provided the pathways employed by users to navigate among the complex array of related bibliographic records. Traditionally, this linkage has been accomplished through the combined efforts of the cataloging code—which governs the descriptive elements and access points of the bibliographic record—and the structure and arrangement of the bibliographic records in the catalog or database.⁶ Thus, in order to fulfill the collocating function of the catalog, the identification and linkage of bibliographic records for related entities are dependent upon the symbiotic relationship between the code and the catalog structure. This dependence continues today, and any discussion of linkages used to express bibliographic relationships must consider both the role of the cataloging code and the structural environment of the catalog.

Dual Function of Linkages

While implying connectivity, the *linkage* frequently serves two distinct functions in the expression of bibliographic relationships. A linkage helps to *identify* a specific or potential relationship among bibliographic entities, and it also serves to *link* the bibliographic records for the related entities. Not all current linkages successfully accomplish both functions, placing the burden upon the user to infer relationships from the combination of information presented in the bibliographic record and the proximity of records in the catalog or database display.

Linkages are divided into two categories: explicit or implicit. Explicit linkages can be either directional links that provide instructions to lead searchers from one heading to another (e.g., cross references), or mechanical links made by a computer that automatically identify related records in some way for the searcher. Explicit linkages serve the linking function well. For example, a directional name-uniform title *see* reference that leads the user from the heading for a part of a work to the heading for the whole work, explicitly links the part to the whole. The user, however, must infer the type of relationship—in this case a whole-part relationship—from the information presented in the reference heading, for the relationship is not explicitly identified. It is the cataloging code that governs this reference linkage.

Mechanical links offer the opportunity to provide both explicit identification of the relationship type and explicit linking of the related records. For example, the 76x-78x linking entry fields in the MARC format, used primarily for serials, may be coded to identify the nature of the relationship—e.g., preceding or succeeding entry—and provide the means for a direct link between the two records. In this case the cataloging code prescribes the information necessary to identify the details of the relationship, but the encoded record structure identifies the specific relationship type and the database structure enables the direct linkage.

Implicit linkages on the other hand, are not specifically identified as links in the catalog; they are usually achieved by one of two methods. Either they can be indicated through collocation, when records with the same heading are placed adjacent to one another in a file or display (*entry level linkage*); or, they can be provided on a bibliographic record through a specific reference or citation (usually in a note) that is not directly linked to another record (*data level linkage*). In the case of implicit links achieved through collocation, the linking function is served indirectly by the structure of the catalog, for the user must infer that some relationship exists based on record proximity. The identifying function is rarely served by collocation, however, for proximity cannot indicate the specific nature of a relationship.

Most implicit links that appear as specific citations on the bibliographic record—e.g., a bibliographic history note citing an earlier edition—are in fact more of an *identifier* of the relationship than an actual link. The potential exists, however, for computers to convert these implicit identifiers and links into explicit linkages. A variety of options are available including expanding the number of linking fields in the MARC record, linking relationship statements in a relational database structure, or by standardized mark-up language encoding in a hypertext environment.

It is the function of the cataloging code to ensure that relationships are identified and related records linked by prescribing rules to create bibliographic records that contain the access points and identifying data necessary for the task. The information required will vary depending upon the relationship type. It is also the function of the cataloging code to define the level at which the entity is described—*work, expression, manifestation, or item*—and to prescribe whether descriptions for the various levels will be contained in one bibliographic record (as is generally the case now), or whether separate records should be created for each level. The task of creating the explicit linkages, and in many cases the explicit identifiers, is allocated to the structure and encoding of the bibliographic record, and the design of the catalog database.

Current Linkage Systems

Two research studies on bibliographic relationships included a detailed analysis of the types of linkages currently used to express relationships in catalogs. Tillett based her linkage analysis on the cataloging codes, and identified the following linking devices for each relationship category:

Equivalence—Dash entry; notes; uniform title (Tillett, 1987, 42);

Derivative—References; dash entries for added editions; edition statements; notes; uniform titles; cross references; subject headings; common main entries; filing titles; added entries (Tillett, 1987, 56);

Descriptive—Notes; common main entry; added entries; subject entries (Tillett, 1987, 58);

Whole-Part—Contents notes; analytical entries; added entries; multilevel description; dash entries; uniform titles; explanatory references (Tillett, 1987, 71);

Accompanying—Addition to physical description; notes; dash entries; multilevel description; separate record with linking notes (Tillett, 1987, 77);

Sequential—Notes; added entries; uniform titles (Tillett, 1987, 82);

Shared Characteristic—Same access point; language; publisher; date (Tillett, 1987, 83).

In a later article, Tillett provided an history and in-depth analysis of linking devices, grouped into three primary categories: catalog entries, uniform titles, and other devices incorporating linking information (Tillett, 1992b).

Vellucci's research, which was based on literary warrant rather than code analysis, found a similar distribution of linkage types among relationship categories, but grouped the linkages found for musical bibliographic entities into the following four categories:

References—See references; see also references; explanatory references;

Access points—Name-title; uniform title; name; title; series title; subject headings; dash entries;

Notes—Contents; accompanying material; form and medium of performance; language; edition and history; “with;” relationship to reproduction; physical description; responsibility; local holdings; summary; notation; other formats; series; and

Other information found on the bibliographic record—Title and statement of responsibility; edition statement; publication statement; physical description; series statement; musical presentation statement; MARC 040 field; MARC holdings data (Vellucci, 1995, 312-314).

The findings of these two studies show that while the primary linking devices are various access points, the information necessary to identify and link bibliographic relationships is embedded throughout the entire bibliographic record including the ISBD areas of description. The lack of explicit relationship identifiers and direct linkages often forces the user to defer to other data on the bibliographic record to deduce relationship information. For example, name-title entry collocation may alert the user that two records are related in some way because of proximity, but the specific nature of the relationship—in the case of a successive derivative relationship for example—may only be inferred by comparison of information found in the edition and publication statements of the two bibliographic descriptions.

Linkage Types and Associated Relationships

References, which are separate from bibliographic records and consistently provide the most explicit connectivity of all linkage types (other than direct mechanical links), are the only linkages that operate solely at the higher entity level, connecting *works* and/or *expressions*. All other linkages, even when linking *works* and/or *expressions* via access points, are also linking at the lower entity level, for they are connecting bibliographic records for specific *manifestations* or *items*. This, of course, is due to the underlying principle of the present code to “describe the item in hand,” which limits bibliographic records to the description of lower level (i.e., physical) entities. As the IFLA study explained, however, the data on the bibliographic record is incorporative of several entity levels in that it concatenates “attributes of a particular *manifestation*, with the attributes of the *expression* that is embodied in that *manifestation* and

with the attributes of the *work* that is realized through that *expression*” (IFLA, 1996, 43). In the current linkage system, therefore, references provide solely higher level entity linkage, while bibliographic records—which include both higher and lower level entity descriptive information—are linking at the higher and lower entity level, simultaneously.

When examining the specific linkages associated with a given relationship type, it should be remembered that the findings of both Tillett and Vellucci are dependent on the linkage treatment prescribed by the cataloging code. Tillett’s findings resulted from a direct analysis of the codes’ treatment of linkages, while Vellucci’s findings were based on the examination of bibliographic records that were created according to the conventions of the cataloging codes. It is possible, therefore, that in a theoretical model, other linkage-relationship associations may operate as well.

Both Tillett and Vellucci found reference linkages to be associated exclusively with whole-part and derivative relationships.⁷ This, no doubt, is because the cataloging rules do not make specific provision for references for other types of bibliographic relationships. Examination of the IFLA higher level relationship structures confirms that the vast majority of relationships at the higher entity level are indeed whole-part and derivative; however, the IFLA analysis also identified sequential and accompanying relationship possibilities at this level. Although not currently used as linkages for these two relationship types, the rules do allow for references to be made instead of added entries (26.6A), and for references to be made for relationships other than whole-part (26.4C1), although the example given in the code is for the more commonly found derivative relationship. Depending on the structural design of the catalog, the expanded use of references in lieu of added entries for linking higher level entities, may merit further exploration. In this regard, some cataloging theorists believe that many aspects of collocation have been abandoned by the substitution of added entries for references, and that the user would be better served by a return to references.⁸ This view warrants further discussion, and should be examined in relation to its impact within various catalog environments.

Access point linkages, including main, added, and analytical entries, are probably the most common form of linkage used in today’s catalog. They provide implicit links for most relationships by collocating and displaying common headings together in the catalog. Identification of the relationship type is implicit, and depends on the user’s ability to infer relationship information from record proximity and cataloging conventions. Access points in one form or another are found as linkages for every relationship type. Although name-uniform title access points are used to identify *works* and *expressions* that are abstract in nature, the linkage does not exist solely at the abstract level, since, as explained earlier, these access points occur as part of bibliographic records for specific *manifestations* or *items*.

Among the several types of access points, the name-uniform title *citation* access point, which is used to identify the *work* and is presently bound to the concept of main entry, is common for all types of relationships, and reflects the current catalog structure’s reliance on entry level linking. Even when another linkage is used to identify the relationship type, the function of actually linking the related records usually falls to the *citation* access point.

Within the citation construct, the uniform title plays a major role in identifying and linking related records. The *AACR2* glossary defines a uniform title as “the particular title by which a work that has appeared under varying titles is to be identified for cataloging purposes” (*AACR2*, 1988, 624), providing the clue as to why it is so heavily used among derivative relationships. When a work appears under varying titles, it implies that derivative manifestations of the work exist. The uniform title by definition then, is the means of collocating derivative entities in the catalog. Additionally, it is used to collocate equivalent entities, to identify the whole-part relationship, and to link a part to its whole in the catalog. The potential importance of the uniform title in a variety of future catalog structures will be discussed later in the section “New Catalog Environments.”

Even though uniform titles play a major role in the expression of bibliographic relationships, there are certain situations in which the cataloging rules fail to provide even this implicit level of linkage. For example, the rules state that a uniform title should not be used for a revised edition of a work when it is in the same language as the original (25.2B). In cases where the title has changed in the revised edition, the related works will not be collocated in the catalog; and while a note on the record for the revised edition should direct the user to the previous edition, there is no reciprocal link from the earlier edition that will lead the user to the later edition. Yee provides several detailed examples of this problem in her discussion of the work, the user, and the objectives of the catalog (Yee, 1994b, 19-21). As descriptions of electronic documents become a larger part of the catalog, the difficulties of describing and relating numerous versions of these entities will magnify the predicament.

Additional problems arise because the rules, with their focus on item level cataloging, do not require normalization of the heading for a work. The rules for construction of a uniform title have limited application (as seen above), and their application is not required even when a particular condition for use may be met. General rule 25.1A states that “Although the rules in this chapter are stated as instructions, apply them according to the policy of the cataloging agency.” This creates inconsistencies both among catalogs and within a specific catalog, weakening the collocation of records for related material. If the focus of descriptive cataloging shifts from the item to the work level, consistent application of the rules for uniform titles will be necessary.

Other access points that appear with regularity as linkages include series added entries for whole-part and sequential relationships; title entries for derivative collections and media materials; subject headings, primarily for descriptive relationships, but also used for derivative relationship; and name access points for shared characteristics.

While the strength of access points is in providing a level of implicit linkage in the catalog rather than identifying relationships, the notes linkage category is generally more successful at identifying relationships than linking related records (with the exception of the MARC linking entry fields previously discussed). The AACR call for the use of notes to provide information not contained in other areas of the bibliographic description. While the function of notes extends well beyond their use as relationship identifiers or links, they are among the more important linkages for bibliographic relationships. This is due to their overall function within the bibliographic record, i.e., as supplemental information to the bibliographic description. Since the main body of the current bibliographic record is designed for description of the item, the information necessary to identify related entities is provided within the supplemental information in notes. Notes with a formal construct, such as those prescribed for contents and summary notes, are more explicit in identifying the type of relationship.

Among the wide variety of notes prescribed by the AACR, those most commonly used to identify bibliographic relationships include the contents note, frequently used for whole-part relationships; accompanying material notes, used for both accompanying and derivative relationships; language, bibliographic history, and statement of responsibility notes, most often associated with derivative relationships; reproduction detail notes, primarily used to indicate equivalence relationships; and summary notes that express the descriptive relationship. These are not exclusive associations, however, and most note types are associated with several relationship categories.

Many changes needed in the notes linkage category to enhance the treatment of bibliographic relationships depend more on improvements to the record and catalog structures than on code revision. Nevertheless, one problem area that is also related to the code is the contents note, an important identifier for whole-part relationships and a popular bibliographic data element for users.⁹ Currently, access to information in most note fields depends on the success of keyword

searching, which is hindered to some extent by the lack of authority control for names and titles appearing in such notes. Names with variant forms and works requiring uniform titles create collocation problems for catalog searches in fields without authority control. The problem arises because the cataloging rules call for presenting contents notes data in the form in which it is found on the item being cataloged (2.7B18).¹⁰ Thus, in most current catalogs, the use of notes to fulfill the linking function depends largely on the keyword search capabilities of individual OPACs using a hit or miss combination of controlled and uncontrolled data. If the AACR moved away from the “item in hand” principle, and allowed authority data to replace data transcribed from the item in the contents note area, this would improve retrieval for keyword searches in the short term, and would pave the way for direct linking of bibliographic records at the data level, or direct linking of bibliographic records and authority records at the data level. If this concept were applied to name and title information presented in other notes areas, such as bibliographic history or statement of responsibility notes, it would improve the identification and linkage for several types of bibliographic relationships.

In addition, with the move toward minimal level cataloging and core bibliographic records, the cataloging community should be aware of how the lack of certain notes will affect the identification and linkage of records for related materials. For example, without the bibliographic history note citation to an earlier edition, the identification and linking of revised editions with title changes in current catalog structures would be severely curtailed.

Several of the ISBD areas of description provide the necessary detail to identify relationships when other means are lacking, but generally cannot provide explicit linkage to related records. One frequent use of ISBD data is the addition of information to the physical description to denote an accompanying relationship. Other uses include edition statements that are often used to alert the user that other editions exist without specifically identify those other editions; and series statements, that tell the user of a sequential and whole-part relationship without creating the linkage provided by a series added entry.

In summary, the primary linkage categories currently in use to express bibliographic relationships in catalogs include references, access points, and notes, with additional information provided by other ISBD areas of description. Only references and a few direct linking notes provide any explicit linkage of related bibliographic records, and with the possible exception of formally structured notes, few currently used linking devices clearly identify the type of relationship. The two processes necessary for the successful treatment of bibliographic relationships—identification and linkage—are accomplished through the combined efforts of the cataloging rules and the bibliographic record and catalog structures. Identification is primarily dependent on information prescribed by the code, while linkage is usually dependent on record structure and catalog design. Because of the dependence on catalog design, most bibliographic relationships currently rely upon implicit entry level linking through record or citation display proximity, implicit relationship identification through descriptive information in the bibliographic record, and the user’s ability to infer relationship meaning from a combination of the two.

Relationships Important to Catalog Users

It is clear from this discussion of linkages that the lack of explicit identifiers and linkages places a great deal of responsibility for determining relationships on the catalog user and his or her ability to infer relationships from record proximity and data in the record. Deducing the type of bibliographic relationship based on the type of linking device used is made more difficult because the cataloging rules often prescribe the use of one device to explicate several different relationship types, while at the same time several different linking devices are used to identify the same relationship. Consequently, there is not a 1-to-1 correlation between linkage type and relationship.

An additional confusion for the user is that linking devices used to show relationships often have other uses as well in bibliographic description.

This may explain the problems faced by users who need to understand a given relationship, but it does not identify the catalog users' needs regarding bibliographic relationships. In order to redesign catalog structures and provide cataloging rules that create effective tools for access and description, the relationships that are most important to the user must be identified.

User Studies

In a paper presented at a Seminar on Bibliographic Records, Svenonius discussed how the data elements on bibliographic records fall into two categories: "those that describe the entity in hand and those that relate the entity to other entities" (Svenonius, 1992, 7-8). The former she termed descriptive elements and the latter were referred to as organizing elements. The organizing elements included the normalized forms of names and titles, including uniform titles, main entries, and added entries. Svenonius stated that the purpose of these organizing elements "is to convey information about bibliographic relationships and in so doing to organize the catalogue" (Svenonius, 1992, 8). These two categories of data elements have each been the focus of catalog user studies, but unfortunately not within the context defined by Svenonius.

Although dozens of studies have examined various aspects of catalog use,¹¹ the resulting empirical data generally is inadequate to clarify user needs in terms of bibliographic relationships. Methodological problems have contributed to this lack of applicable user data. With the increasing number of OPACs, transaction log analysis has become the data collection method of choice, and while this does provide an unobtrusive window from which to examine catalog use activity, it does not answer questions about whether the user is searching for a work (represented by more than one record) or an item (represented by one specific record), or if and how the data elements in the bibliographic record are used to identify a specific relationship. Another problem is the users' lack of a vocabulary to discuss their information needs, their search processes, and the bibliographic data used (Yee, 1991, 90). This creates problems for methodologies that employ interview techniques, for a user unfamiliar with the concept of bibliographic relationships, the means used to identify relationships, and the technical terms used to describe the parts of a bibliographic record, will find it difficult to converse about such needs and data uses.

Many studies have investigated how users search the catalog, focusing on the access points employed. A large portion of these studies have examined subject access, which is not part of the descriptive cataloging process, and therefore, only marginally associated with this discussion of bibliographic relationships. When subject access is omitted, almost all studies found that the most used access points were title and author,¹² both of which have been identified as linkages for bibliographic relationships. For several reasons, however, caution must be exercised when interpreting these results to infer user need for the bibliographic relationships associated with these access points. First, most user studies identify author and title access point searches as "known item" searches, assuming that a searcher uses these access points when a specific *item* is known and sought. This would preclude the need to identify any related materials. None has taken into account the possibility that the user was actually performing a *work* search, to determine what editions or versions of a work might exist and then use the collocated list of related records to select an appropriate item; nor has any study provided data on the proportion of "known item" searches that retrieved multiple records for one work. Once categorized as a known item search, most user studies did not pursue further the user's motives and goals in performing the search. Such a follow-up might have discovered that the searcher was actually employing the "second objective" of the catalog (collocating), to identify specific relationships of importance, or in Svenonius' terms, employing the organizing elements of the bibliographic record to convey information about bibliographic relationships. As it stands, it can only be conjectured that some

proportion of these known item searches were work searches that necessitated identification of relationships among entities represented by the retrieved record set.

A second problem with studies that focus on access points is that several relationship types use the same access points for linkage. Even if it were assumed that an author, title, or author/title search was conducted to view the record set for a work, rather than a known item, it would be impossible to determine which specific relationship was sought by the user without an interview. A third problem with many user studies is the lack of differentiation between title proper and uniform title. Such a distinction might indicate whether the search was for an item or a work, and might indicate the user's need for specific relationships to be identified. This is a particularly thorny problem to solve because the user might not understand the distinction between the two types of titles, and the indexing of many catalogs will retrieve data from both title proper and uniform title fields when the titles are identical or similar, depending on the search method employed.

It is interesting to note that the form of the catalog has affected how its use is studied and what aspects of use are examined. While the focus of more recent research has been on access points and user searching as determined by transaction log analysis, earlier user studies—primarily of card catalogs—focused more on the use of various elements of bibliographic data in the catalog record, employing survey, interview, and observation methodologies. No studies, however, have provided data on the use of data elements in the catalog record within the context of bibliographic relationships. Palmer summarized 22 catalog use studies prior to his own research in 1972 (Palmer, 1972), and discovered that overall the most heavily used data elements were author's name, title, call number, subject headings, and date of publication. The moderately used elements included place of publication, publisher's name, edition statement, and contents note. All remaining elements were used infrequently. Palmer's own study largely supported these findings. Among the few recent studies that examined the bibliographic record, Hufford's examination of reference librarians' use of data elements indicated that 20.6% of all data elements accounted for 96.6% of the total uses of all elements (Hufford, 1991, 59), with title proper, author, location information, call number, serial chronological designation, date of publication, and serial alpha/numeric designation accounting for the most heavy use. The study by Lundgren and Simpson, which examined faculty opinion on the usefulness of specific data elements of the bibliographic record, also supported similar findings (Lundgren and Simpson, 1997). Those elements rated as the highest degree of usefulness included title, primary author, date, subject, other authors, and series; those rated as a moderately high degree of usefulness included summary notes, contents notes, standard numbers, publisher, pagination, and related titles. Since these studies did not indicate how the information was used, one can only guess that some portion of these data elements were employed to assist in the identification of a relationship.

Additional problems exist with many user studies. Most examined a limited user population (i.e., staff, faculty, academic library users, etc.), providing data that is not generalizable to a larger population, or to users within specific disciplines who may have different needs. For example, Hufford's study of reference staff use concluded that "accompanying material" data were not used very frequently; however, it is likely that a similar study in a music library would find this data element heavily used to identify a musical score with accompanying parts.

Studies in the area of information retrieval have focused on the cognitive and behavioral aspects of catalog use, considering various factors that might affect the search, and trying to determine the users' criteria for successful catalog searches.¹³ While the methodology employed in these studies would be helpful in determining a user's need for identification and linking of bibliographic relationships, none has examined that particular aspect of catalog use.

What then, if anything, can be learned about bibliographic relationships from the user research conducted thus far? Two of the most important linkages for derivative, whole-part, and

equivalence relationships—author and title—are among the most heavily used access points. Of the data elements discovered to be heavily or moderately used, the date of publication, place of publication, publisher's name, and edition statements might be used primarily to demonstrate derivative relationships; contents note use would mainly indicate a need to identify whole-part relationships; and serial data elements would associate with sequential and whole-part relationships. It is also likely that the importance of various data elements and relationships will vary depending on the discipline, the type of library, and the type of user. Individual testimony, for example, has indicated that historical research scholars rely heavily on data elements that relate editions, relate and distinguish various states and versions of a work, link original works with their facsimiles, and link parts of a work with the whole (Snyder, 1992).

If these relationship linkages and identifiers were ranked in order of importance based upon the data gathered in these user studies, it would appear that those relationships that served an access function would rank highest, followed by those relationships serving a distinguishing or differentiating function, and those that identify the nature of a relationship. All of this is highly speculative, however, and is based on data culled from studies that were not designed to investigate bibliographic relationships.

IFLA Study

The final portion of the IFLA report attempted to place the bibliographic record within the context of user needs by mapping each entity attribute and relationship to the user tasks that they support (IFLA, 1996, 63). Although the source of the defined user tasks was not stated, they are clearly based on the functions of the catalog rather than any specific empirical user studies. These functions, identified by the IFLA group, included the following four generic user tasks: to find, to identify, to select, and to obtain. Each attribute and relationship was then assessed a value based on the study group's decision on its importance for supporting each of these four user tasks (IFLA, 1996, 64). In many cases, information named as entity attributes in the IFLA study currently serves as a relationship identifier, for it provides the information necessary to distinguish between two entities, thus enabling the nature of the relationship to be understood.

The report assigned the highest value to those relationships that provided information needed for access; information needed to identify, distinguish, or differentiate bibliographic entities; or information about restricted use or access. Within these categories, relationships that were applicable to the largest number of records were assigned the highest value. Additionally, dependent relationships that linked referential entities and component parts were assigned the highest value. Thus, the IFLA study ranked the importance to the user of a given relationship first on the function served, then on its frequency of occurrence, and finally on the nature of the relationship.

This IFLA document is the first to provide a set of criteria to rank the importance of bibliographic relationships to users, and as such must be commended. One potential problem with these criteria, however, is the high ranking associated with frequency, for this creates problems for material that does not bulk large in the catalog. Research has shown that certain relationships are discipline or format specific, and even though a relationship may be important primarily to a subgroup of material within a specific discipline or in a specific format, it may still be vital to the user requiring that material. Thus, frequency of occurrence is a less than ideal means of ranking relationship importance to the user, if it means that lower frequency relationships will receive less than optimum treatment in the catalog. This in effect would create a hierarchy of preferred users based on discipline. It is possible, however, that entities with low frequency relationships will also exhibit a relationship characteristic that would override the lower ranking. For example, an accompanying relationship occurring with lower frequency might still be ranked of high importance because of a dependent relationship to another entity. Perhaps, then, these ranking

criteria might serve the user better if frequency were a secondary criterion, with function and the nature of the relationship taking precedence in rank. It is also noteworthy that the IFLA user criteria—function, nature of the relationship, and frequency—reflect the interpreted findings of the user studies. It would be preferable, however, if these criteria could be verified by user studies designed specifically to test the importance of bibliographic relationships, rather than relying on inference and speculation.

Relationships Important to Authority Record Users

Closely allied to the conceptual model of bibliographic relationships is the authority control process as manifested in the authority record. Based on the choice and form of entry as prescribed by the AACR, authority control facilitates the concept of one authoritative form of a name or title, with variant forms linked to the authorized form through references. Both the authorized form and its variants are recorded in the authority record. Thus, the authority record controls two important linkage categories for bibliographic relationship: access points and references. As part of the descriptive cataloging process authority records are associated with two categories of relationships: *bibliographic relationships*, between two or more bibliographic entities; and *name relationships*,¹⁴ between two or more names or forms of a name. While name relationships are important because names are an attribute of most bibliographic entities, authority records for names alone are not the nucleus of concern for bibliographic relationships.

Bibliographic relationships are primarily associated with authority records for *works*. These work records are closely tied to the concept of main entry as the citation heading for the work, and frequently contain name and uniform title components—or a uniform title alone—depending on the origin of the work and the cataloging code. The uniform title is an inextricable part of work authority records, as the very existence of an authority record for a work currently depends in large part on that work's need for a uniform title as determined by the cataloging code and library policy. But just as uniform titles are not created and used for all works, authority records have not been not created for all works, even those requiring uniform titles. For example, Vellucci's study of music bibliographic relationships found work authority records for only 26.1% of the musical works in her sample (Vellucci, 1995, 236), even though 83.3% of the sample had a uniform title, and 62.9% of the sample used the uniform title as a linking device for a bibliographic relationship (Vellucci, 1995, 245).¹⁵ Theorists have predicted an expanded role for the uniform title in the online catalogs of the future due to its crucial function as a work identifier (Gorman, 1979; Wilson, 1983; Vellucci, 1990; Wainwright, 1991; and Heaney, 1995), for in order to develop a fully-linked work-based catalog environment, authority records must be created for all works, which means that uniform titles must be constructed and applied consistently for all works.

Other than Vellucci's research, few empirical studies have examined the use of authority records within the context of bibliographic relationships, although a considerable body of research exists on the authority control of name relationships. An early study by Smiraglia that examined the use of uniform titles for music as one aspect of derivative relationships (Smiraglia, 1989), showed that virtually his entire sample yielded multiple manifestations, and that the majority of these had titles proper different from that of the first edition of the work. From these results Smiraglia concluded that uniform titles were a necessary part of the description of musical works and were needed to serve as authority controlled collocating devices. He also cited the need for more references as part of that authority control.

In reporting her replication of Taylor's classic study of the value of authority files for name searching in an online catalog (Taylor, 1984), Bangalore reiterated that "the authority file is not only for the benefit of users, but also acts as a resource for library staff in technical as well as public services" (Bangalore, 1995, 92). Name authority records are an important time saver for catalogers and, therefore, are useful even when references are not required for a given name.

While discussed within the context of name authority records only, these views could be extended to authority records for works and their usefulness for the identification and demonstration of bibliographic relationships.

Authority records for works are associated most frequently with four types of bibliographic relationships—derivative, whole-part, equivalence, and sequential. These associations stem in part from the *AACR2R* general rule (25.1A) that defines the use of uniform titles, i.e., to identify the work, to differentiate works with similar or like titles, and to collocate works (*AACR2R*, 1993). As discussed earlier, name/uniform title references—determined by chapter 26 of *AACR2R* and controlled by authority records—serve as linking devices for both derivative and whole-part relationships. When included in bibliographic records, the name/uniform title citation headings controlled by authority records help identify relationships and link the records for related entities. Thus, several of the component parts of the work authority record—including names, uniform titles, and references—play an important role in the expression of bibliographic relationships in the catalog.

Recently there has been growing support to replace *authority records* with *access control records*, a change that reflects a shift in concept as well as terminology. Begun by Gorman twenty years ago with the idea of *information packages* as an alternative to the concept of main entry (Gorman, 1979), the term *access control* was first applied by Tillett a decade later (Tillett, 1990). Since then the body of literature on access control has continued to expand, as interest in work-based cataloging and direct-linkage databases has increased (Tillett, 1989b; Gorman, 1992; Tillett, 1990; Tillett, 1995; Heaney, 1995; Tillett, 1996; Barnhart, 1996).

Access control records were described by Barnhart as the next generation of authority records. These records remove the notion of “authority” and link variant forms of names and titles without declaring any one the “authorized” form. One form may be selected as the default form for display, but this can be varied by the user or the library. In a direct-linkage database structure, these access control records would be connected to bibliographic records—which would collocate all manifestations of a work—and to other related access control records—which would collocate related works. The access control record would become the locus for all information obtaining to the work, and could be expanded to include much information that is presently encoded in the bibliographic record for the manifestation (Barnhart, 1996).

The use of access control records could greatly enhance the demonstration of bibliographic relationships in the catalog; however, to be fully implemented certain inconsistencies in the cataloging code and its application for different formats of material must be addressed. For example, the previously discussed problem of revised editions of a work must be resolved so that there is a bi-directional link between the old and new editions. This type of linkage could be accomplished in a variety of ways through the use of relator codes and by linking access control records with each other and with the appropriate bibliographic records.

Another problem exists with the varying treatment of translations for non-music and music materials. As Barnhart pointed out, the *AACR2R* rule for references from variant titles (26.4B1) states that “In the case of translated titles, refer to the uniform title *and the appropriate language subheading...*” (*AACR2R*, 1988, 558, emphasis added). This rule, in effect, results in the creation of separate authority records—one for the original work and one for each language into which that work was translated, thus scattering information about the work among several records. For example, Thomas Mann’s *Magic Mountain* has the following two authority records:

```
010      n 79077983
100 10   Mann, Thomas, $d 1875-1955. $t Zauberberg

010      n 94017974
```

100 10 Mann, Thomas, \$d 1875-1955. \$t Zaubenberg. \$l English
400 10 Mann, Thomas, \$d 1875-1955. \$t Magic Mountain

If these name/uniform title authority records are considered *work* records, then this practice essentially treats the translation as a new work with its own work authority record. While some might agree with this idea for semantic reasons, for purposes of entry the code has adopted the view that they are the same work and the translation should be entered under the same heading as the original (21.14A).

A Library of Congress Music Cataloging Decision (MCD) for rule 26.4B1 allows music catalogers to refer to an authority record for the original work that contains a uniform title without the addition of elements that would be used for derivative manifestations in most cases.¹⁶ This interpretation of the rule allows music catalogers to create one authority record for the work that includes references from the variant forms of the title in many other languages. For example, the following is the work authority record for Brahms' *Zigeunerlieder*:

010 n 82040719
100 10 Brahms, Johannes, \$d 1833-1897. \$t Zigeunerlieder
400 10 Brahms, Johannes, \$d 1833-1897. \$t Gypsy songs
400 10 Brahms, Johannes, \$d 1833-1897. \$t Chants tsiganes

When used on a bibliographic record, the language element \$l is added as appropriate to the uniform title for the original work. This practice allows one work authority record to be created that collocates information appropriate to that work. Although this discrepancy is between the rules and an interpretation of the rules, rather than within the rules themselves, the impact on the creation of authority records and the catalog remains the same.

One final example of inconsistency in the rules to be discussed here is the treatment of entry for whole-part relationships. Rule 25.6A1 states "If a separately catalogued part of a work has a title of its own, use the title of the part by itself as the uniform title. Make a *see* reference from the heading for the whole work and the title of the part as a subheading of the title of the whole work" (*AACR2R*, 1988, 491). The separate part is entered as a subheading of the title of the whole work only if the part does not have a distinctive title (Rule 25.6A2). In a work-based environment, this practice essentially treats the part as a separate work with its own independent heading and authority record; and the determination of that separateness is not based upon the nature of the work, but rather whether or not it has a distinctive title. Special types of material are treated differently, however, and for both sacred scriptures and music, parts of a work with either distinctive or nondistinctive titles are entered under the heading for the whole work, with the title of the part entered as a subheading of the title of the whole work (rules 25.18A1, et al.; 25.32). The end result of these discrepancies is that for some materials there are authority records with the title of the part as the authorized heading and the whole-part title construct as the reference, and with other authority records the opposite is true. The authority records below illustrate this:

010 n 82257989
100 10 Dickens, Charles, \$d 1812-1870. \$t Bagman's story
400 10 Dickens, Charles, \$d 1812-1870. \$t Pickwick papers. \$p Bagman's story

010 n 86130956
100 10 Bach, Johann Sebastian, \$d 1685-1750. \$t Matthäuspassion. \$p Komm, süßes Kreuz
400 10 Bach, Johann Sebastian, \$d 1685-1750. \$t Komm, süßes Kreuz

In the catalog, the bibliographic records for the work and all its separate parts will be collocated for some material, while for other materials the user will retrieve a mixture of bibliographic records and references, but the bibliographic records for the parts will not collocate with those for the whole work. A consistent approach to the entry of parts of works would make the catalogers work easier and bring uniformity to the treatment of whole-part relationships in the catalog for the user.

In summary, authority records for *works* are a critical part of the bibliographic relationship environment. While little empirical research exists that verifies the importance to users of authority records for the control of bibliographic relationships, observation indicates their importance for maintaining citation headings for works and providing the reference structure necessary for certain types of bibliographic relationships. These authority records are dependent upon the cataloging code to prescribe guidelines for constructing the citation headings—an important part of which is the uniform title—and the necessary references. The present creation of name/uniform title authority records is erratic at best, and does not represent the systematic process that is necessary to support a work-based catalog structure capable of treating bibliographic relationships with greater precision.

Underlying most of the problems discussed in this section is the code's lack of a conceptual foundation for describing a work and creating work authority records, a deficiency that stems in part from the focus on item level cataloging and in part from the code's lack of a clear definition for a work. This results in inconsistencies in the construction and application of uniform titles and their authority records. It also leads to a proliferation of specialized rules and rule interpretations for materials that exhibit numerous relationships and are, therefore, more dependent on the concept of the work for their description, access, and expression of relationships.

Improved Methods for Expressing Bibliographic Relationships in the Online Environment

By now it is evident that while the rules often provide for the descriptive detail and access points necessary to identify a particular relationship, linkage—and in many cases explicit identification—is dependent upon the record and catalog structures. For better or worse, the MARC communications format is the most prominent record structure in use, and is the structure upon which most of today's OPACs are based. Since vast amounts of money have been invested in the MARC format in terms of the huge number of encoded records and the various library systems and utilities designed specifically for compatibility with the present formats, the *Realpolitik* of the situation suggests that the MARC structure will not be discarded lightly, if at all. Consequently, discussions of catalog improvements are often within the context of revamping the MARC record structure to accommodate better a direct-linkage database environment, rather than abandoning its use altogether.

The MARC Environment

A good starting point for discussion of bibliographic relationships within the MARC environment is Leazer's analysis of the USMARC formats for bibliographic, authority, and holdings data, conducted to determine the degree to which the structure adheres to a conceptual design schema as defined by database design theory (Leazer, 1992). With a view towards supporting future catalogs capable of formally expressing bibliographic relationships, Leazer analyzed the compactness, efficiency, and expressiveness of the MARC record format. He discovered a high degree of redundancy among the data elements and their functions, with 26 specific fields to handle bibliographic relationships, and a total of 22 fields designed to link related records. Much of the overall redundancy was due to the use of the MARC format within an

antiquated linear computer environment. For example, much redundancy was caused by the expression of the same data in both encoded and narrative form, which could be eliminated by exploiting the computer's ability to store data in one format and display it in another. In other instances the same data was repeated in two different linked records, when a direct-linkage database structure could have connected the records without repeating the data. Both of these shortcomings have implications for the explicit expression of bibliographic relationships. For example, a relationship type could be identified by encoded data that could then be displayed as a textual label; or records for related material could be linked using a record control number and coded to display textual information from a specific field without rekeying the data. Leazer concluded that a new conceptual schema for the organization and structure of bibliographic databases was needed in addition to a reconceptualization of the MARC format. This dual theme of revamping MARC and developing a new database structure in order to improve the catalog's relational capabilities has become the ground bass for many articles on this topic.

As Leazer noted, the current MARC record structure already accommodates certain bibliographic relationships and record linkages; with revision, it is capable of providing the more sophisticated record linkages, data linkages, and explicit identifiers that are needed to create more formal expressions of bibliographic relationships in the catalog. MARC revision can be discussed on a continuum ranging from slight modifications, such as increasing the number of specialized note fields that deal with bibliographic relationships and expanding their record linking capabilities, to creating an entirely new MARC framework designed for direct-linkage database systems and hypermedia structures.

To this end, Tillett proposed a "slow evolution" in which current MARC fields could be mapped into a new MARC structure (Tillett, 1992a), with traditional linking devices such as notes, combined with nontraditional next generation hypertext links that would not only connect records, but identify relationships. In her vision of the new "MARC III" format, there are no separate access fields, but rather text within the descriptive fields that is intended for access would be "marked" in such a way that it would be connected to the appropriate access control record for that data element. Tillett's "MARC III" structure requires direct links between access control records and bibliographic records, and a holdings format that links item specific information to the appropriate bibliographic record. The entire structure is designed to "provide ways for users of online systems to discover the relationships and interconnections inherent in bibliographic information and those accidental, incidental, or perhaps even mystical relationships that derive from sharing an attribute..." (Tillett, 1992a, 116). Again, this idea represents MARC revision coupled with database restructuring, but Tillett's vision of the new "MARC III" catalog structure is grounded in the belief that the cataloging rules and principles should transcend the format used to communicate or store the bibliographic data. In her view, the cataloging code should be an underlying constant for the creation of cataloging records, and as such should remain separate from the record format (Tillett, 1992a, 117).

Other theorists have advocated a more radical approach, suggesting totally revamped MARC and database structures, and cataloging rules that integrate MARC coding with description and access. Gorman described the current MARC format as a unitary record that contains complex information with few and weak links to other related unitary records (Gorman, 1992, 91). He proposed that this structure be replaced by a system with many records containing simple data, wherein the records could be linked in a variety of ways to create a complex structure expressive of all types of bibliographic relationships. Gorman's view recasts both the cataloging rules and the MARC format into *The HYPERMARC Record Preparation Manual—Bibliographic*, which integrates the rules for description, access, and record formatting into a single document. Significant features of this new manual include the provision for "levels" of description (in terms of analysis), guidelines for linking these descriptions, instructions for creating authority packages, and guidelines for linking authority packages to bibliographic descriptions and other information

packages. Once again, the expression of bibliographic relationships is a key factor in the revision of MARC and the restructuring of the database.

Ultimately, the record structure should provide an economic and efficient way to store bibliographic data, show the relationships among that data, and present the information in a useful way to the catalog user. Revising the MARC format to accomplish this may be the most financially viable path when the extent of existing investment is considered. Web-based bibliographic systems have already found ways of using hypertext mark-up language to link bibliographic records with index entries, and to establish direct links between information in bibliographic records and the electronic documents that they describe. All of this has been accomplished within the framework of the MARC format. Some researchers, however, do not consider this “band-aid” approach to MARC revision the best, or most efficient approach to database design, even when augmented by hypertext mark-up language linking capabilities.

New Catalog Environments

As long as the structure of the catalog remains modeled on the linear card catalog, improvement in the expression of relationships in the online environment will continue in piecemeal fashion, by combining revision of MARC content, codes, and structure, with a hypertext mark-up language. But as Tillett noted, current hypertext mark-up languages are excellent for direct connections but do not explicitly identify relationships (Tillett, 1992a); and as Leazer found, the current MARC format is riddled with redundancy, and would require extensive reconceptualization to become a truly efficient data structure in a relational database. Several researchers, therefore, have begun to explore the possibility of other database models for future catalogs, some of which use a modified MARC format as an interim data structure, while others focus solely on the database structure without reference to MARC.

Heaney used object-oriented modeling to conceptualize a three-tiered catalog design independent of any particular method of implementation (Heaney, 1995). The three-tiered structure consists of the abstract work, the publication, and the copy. Based on the principle that objects can pass on their characteristics to other related objects, the record for each level of the hierarchy consists of bibliographic data pertinent only to the object at that specific level, with data pertinent to objects at other levels associated with the related object by links. The precise nature of the relationship can then be modeled as a link attribute. Thus, unlike the current MARC structure that includes data for all three tiers in each record, there is no redundancy among the database records. This model “involves a fundamental shift away from the AACR2 philosophy of description of, plus access to, physical items,” and requires cataloging rules that focus on the work (Heaney, 1995, 152). Uniform title authority records for works are a critical component of the structure, and act as a means of linking works with their related bibliographical items. Heaney also suggests that relator codes, similar to those already authorized for use in \$4 of MARC name fields, be used to identify a specific relationship or association, and provides examples of how the MARC format could be adapted as an interim structure for the model. Heaney’s object-oriented model is similar to the entity-relationship model presented in the IFLA study.

The importance to the basic database structure of uniform title authority records for works is echoed by Wainwright, who suggests that a consistent means of linking related versions of works through authority records for the primary work is needed (Wainwright, 1991, 13). Wainwright also notes that “the most recent version of UNIMARC allows for pointers to the authority control records for the work, but unless the various types of relationships can be separately tagged, in any sizeable catalogue the user is likely to be confronted with a confusing group of hits for many author or author-title searches” (Wainwright, 1991, 13). Creating a link between records without identifying the precise nature of the relationship solves only half the problem. Thus, the combined

use of specific relationship identifiers and direct linkages to access control records would be one viable method for identifying and linking related records.

Leazer has explored a different catalog model based on a bicameral system of description that contains one file for descriptions of items and one file for descriptions of works (Leazer, 1994). The individual records in both files will be linked to express explicitly three possible types of relationships: work-to-work, item-to-item, and item-to-work. In this system Leazer views all relationships as binary, one-to-one entity relationships, which may be repeated as often as necessary to describe the full array of possible relationships between any two entities. Leazer's model does not address the use of authority or access control records within this bicameral bibliographic system, but he does suggest computer data recognition routines to identify relational information resident in current MARC records.

Because the concept of a relational database design for catalogs has been growing in popularity,¹⁷ Green's seminar class conducted an examination of the full-relational structure for use with bibliographic, authority, holdings, and classification data (Green, 1996). Building on a detailed analysis of the functions of a bibliographic database,¹⁸ a conceptual database was designed based on an entity-relationship model. The study found that the complexity of bibliographic data created database efficiency problems when the normalization process of record deconstruction was applied to create the relational design. Due to the single-value attribute structure of a relation, any multi-value attribute must generate additional relations. With the high number of multi-value attributes for bibliographic data identified by the analysis (almost one-third of the relations), it would be necessary to access dozens of relations to assemble the complete bibliographic information for a specific physical item. It was determined, therefore, that the cleanness of the full-relational database design was immaterial in the face of the processing inefficiencies discovered. The group concluded that a full-relational approach, while semantically sound, resulted in inefficient processing, and that use of a MARC-based data storage format retains efficiency at the expense of semantic integrity (Green, 1996, 213). Green goes on to suggest more efficient alternative structures, including a MARC-based structure with increased relational links between records, and an object-oriented database structure that accommodates multiple values, and, as Heaney noted, has the added advantage of hierarchical characteristic inheritance.

Thus, while it appears that a new catalog structure would enable bibliographic relationships to be identified and linked in a more direct and logical manner than the current MARC format allows, it is still unclear as to which database design would be most efficient. It is likely, therefore, that for the immediate future the MARC structure will be retained—while increasing its linking capabilities—and any major shift in database structures will occur only after further research has been conducted.

Local vs. Global Bibliographic Universe

Until recently, the role of the local catalog has been to provide access to physical items owned by the local library, justifying to some extent, AACR2's emphasis on description of, and access to the physical item. As Wilson predicted, however, the role of the local catalog is expanding to incorporate information about objects neither owned by the local library nor traditionally accessed through the local catalog (Wilson, 1989a), with a resulting emphasis on the retrieval of works rather than physical objects. This shift in focus from a local to a global bibliographic universe has distinct implications for the treatment of bibliographic relationships in catalogs. Three areas in particular deserve further consideration: the impact of network access to, and simultaneous searching of multiple catalogs and databases; the ability to catalog within an international arena; and the effect of hypertext linkages on surrogate/document relationships.

Network technology offers many opportunities to expand the OPAC on various levels by providing remote access to local and union catalogs via local, regional, national, or international networks. Web-based OPACs using the Z39.50 information retrieval protocol are becoming commonplace, and projects are underway to provide simultaneous searching of multiple catalogs via the World-Wide-Web.¹⁹ The extended local catalog presents a paradox for the expression of bibliographic relationships. On the one hand, since the catalog of any one institution contains only a portion of the bibliographic universe, relationships are likely to exist that are not apparent to users of a local OPAC. Thus, for a more accurate picture of any given work and all of its relationships, bibliographic records in the local catalog should be considered as only one part of the broader bibliographic universe. On the other hand, as the ability to perform simultaneous searches of multiple catalogs allows users greater access to the broader bibliographic universe, the complexity of relationships between records will increase, and users are likely to be overwhelmed by large retrieval sets that provide an inadequate means of interpreting the nature of relationships among myriad records.

One solution to this problem is under development by researchers at the University of Bradford, who have created a user friendly PC-based front-end system for manipulating large retrieval sets (Ayres, Nielsen, and Ridley, 1996), and are currently working on a similar web-based system.²⁰ Based on the concept that works often appear in many versions or with complex multi-part relationships, the Bradford OPAC “groups together into sets items that are manifestations of the same work” (Ayres, Nielsen, and Ridley, 1996, 3). A manifestation set is created around a manifestation entry consisting of a uniform title and author(s) for the work—essentially the equivalent of a work access control record. All manifestations of the work are linked under this heading to create the set, and users are then able to sort, search, and display these sets in various ways. The expression of bibliographic relationships is the primary function of this catalog, and therefore, the difficulties encountered during its development process deal directly with the practical problems of creating a global catalog fully expressive of bibliographic relationships. Many difficulties occurred because of inconsistencies and redundancies in the MARC record. Additional obstacles were caused by the division between main and added entries in AACR2 and the MARC record, as well as the lack of analytical records for works. Thus, the Bradford OPAC project provides an example of how the short-comings of AACR2 and the MARC format made the identification and linking of related records problematic.

The growing emphasis on networked access to library catalogs has engendered an increased awareness of the importance of bibliographic relationships for cataloging on an international scale. Tillett’s access control record concept for names and works, which includes all pertinent data for the entity without identifying one authorized form, could become the pivotal data structure upon which the international cataloging community could develop their global catalogs. The access control record could compensate for differences in language and language scripts by allowing all variant forms of names and titles of works to be equally accessible. Additionally, access control records could provide the vertical and horizontal linking structure necessary for any relationship-based database.

The UNIMARC/Authorities format may be an appropriate structural basis for developing an access control record. In a comparison of the USMARC and UNIMARC authorities formats for the purpose of international data exchange, Truitt noted that “Many of the most valuable links that UNIMARC seeks to establish are not presently coded in USMARC records” (Truitt, 1992, 54). Linkages important to bibliographic relationships in an electronic environment include language specification—i.e., language of cataloging, primary and alternate character sets, and script of cataloging—collective uniform title identification, and explicit inter-record links. Truitt attributed the lack of structural links between related records and headings in the USMARC format to a separate manual authority system mentality focused on the local catalog, and posited that the creators of UNIMARC/Authorities had a better understanding of the need for the identification of linkages and relationships because of the complexity of authority control at the

international level. As discovered by the Bradford OPAC project, not only is there a need to improve the bibliographic relationship linking capabilities of UK and USMARC, but greater compatibility is needed among the various national MARC formats to facilitate the creation of new catalogs based on international data transfer.

A final area of consideration in the local versus global bibliographic universe is the effect of hypertext linkages on surrogate/document relationships. In a discussion of methods for organizing electronic resources, Vellucci noted that the transformation of the catalog into a “gateway” for accessing electronic resources has created conceptual problems for librarians by blurring the boundaries between the catalog and the bibliographic universe (Vellucci, 1996, 13). Until now, linkage concerns have focused on connecting two bibliographic records in the catalog, or in a few more advanced systems, connecting related authority and bibliographic records. With the advent of web-based catalogs, however, the linking function has been expanded to connect information in the bibliographic record to the actual related document when it is available in a digitized format. This example of connectivity represents linkage at the *document* level, and raises new issues about relationships between the surrogate record and the document represented by the surrogate. While this type of linkage is currently made by encoding the data with a hypertext mark-up language, the bibliographic record must contain the necessary information in order to make the link. It is the function of the cataloging code to ensure that the required information is contained in the bibliographic record.

The hypertext environment also affords the opportunity to go beyond connections between the bibliographic record and the electronic document. For example, links can be made from the bibliographic record to descriptive information about the document such as an abstract, summary, or online review of the work. In addition to text, these descriptions may be in graphical, video, or aural form, and the descriptions may be linked as separate objects or embedded in the bibliographic record itself. These types of descriptive relationships will become more common as the technology advances, and the next cataloging code must consider the possibility for including such relationship data in the bibliographic record.

In summary, it appears that no matter which specific database structure ultimately is deemed best for future catalogs, certain issues must be addressed in order to develop an efficient bibliographic system capable of serving the global community of the 21st century. Based on information gleaned from the Bradford OPAC project and other research into direct-linkage database structures, bibliographic relationships are the focal point of many important issues. Some concerns primarily deal with the MARC format’s capacity to handle relationships—i.e., its redundancies, structural linkage deficiencies, and inconsistencies among different national structures. At the heart of many other issues, however, lies the cataloging code that determines the content, access points, and philosophical basis for what is described.

General Principles for Bibliographic Relationships in Catalogs

This paper has shown that bibliographic relationships do not exist as an isolated concept, but are integrated into almost every aspect of the descriptive cataloging process, and are critical to the structure and design of catalogs. It is vital, therefore, that the cataloging rules clearly and directly address the expression of bibliographic relationships as part of the underlying principles that guide the cataloging code. These principles of description for bibliographic relationships are designed to support an expanded version of the catalog’s functions, in which the expression of relationships is an integrated goal, the provision of access to bibliographic entities stored locally or remotely is an intended aim, and the description of bibliographic entities at multiple levels is considered useful. In order to fulfill these functions, the following four principles for bibliographic relationships should guide the cataloging code.

1. *The Principle of Relationship Identification.* The bibliographic record should identify all important bibliographic relationships that exist between the entity being cataloged and other entities. These relationships include both independent & dependent relationships. Identification should be bidirectional.
2. *The Principle of Enabling Linkage.* The data elements of the bibliographic record should enable related bibliographic records to be linked, and should permit the bibliographic record to be linked to related documents. To this end, the bibliographic record should provide enough information to identify the relationship and create a linkage. Linkages between bibliographic records should be bidirectional.
3. *The Principle of Multi-level Description.* The cataloging code should provide for the independent description of an entity at several levels, including the abstract work, the physical item, and the specific copy. These hierarchically related descriptions should be linked.
4. *The Principle of Consistency.* The identification and linkage of like bibliographic relationships should be treated in a consistent manner, regardless of physical format. This includes the consistent application and use of uniform titles.

These principles provide a logical direction for the treatment of bibliographic relationships in the cataloging code. They encourage a consistent and uniform approach that would enable all formats of material to treat relationships in the same manner. The application of these principles would facilitate record restructuring and catalog design, expedite the international transfer of bibliographic data, and enable the design of a bibliographic system with specific relationship identifiers and direct linkages. Above all, they would allow the construction of catalogs that provide users with an overall picture of the work sought, its many and varied manifestations, and any works or manifestations of a work that are related in some way. In conclusion, "Cataloging is an art, not a science. No rules can take the place of experience and good judgement, but some of the results of experience may be best indicated by the rules" (Cutter, 1904, 6).

Notes

1. Gorman saw the definition of bibliographic relationships as the basis upon which the catalog file structure depended. He viewed the links between bibliographic records as critical to file design, and called for studies of the universe of bibliographic relationships in order to ensure that these links were both comprehensive and practical (Gorman, 1978, 193).

2. Attig was concerned with the development of structures for communicating information about bibliographic relationships when he expressed his opinion that “Before decisions can be made or data structures designed, ... there must be some solid conceptual analysis of the bibliographic universe” (Attig, 1989, 145). More recently Gorman called for bibliographic system records that contained simple data with articulated, complex, and sophisticated linkages expressive of all the pertinent bibliographic relationships (Gorman, 1992, 92-3).

3. Since the publication of Fidel and Crandall’s article “AACR2 as a Design Schema for Bibliographic Databases” in 1988 (Fidel and Crandall, 1988), interest has increased in developing new catalog structures, primarily based on databases that employ data linking based on relationships. Leazer’s doctoral research was concerned with developing a relational database model to accommodate better the bibliographic relationships associated with works (Leazer, 1993). Heaney discussed the cataloging process in terms of an object-oriented database model for catalogs (Heaney, 1995).

4. To cloud further the issue, manifestations are frequently called versions or editions, and catalogers usually disagree with bibliographers on the definitions of these terms.

5. Note, for example, the variation in subclasses defined by Vellucci in the study of music bibliographic relationships (Vellucci, 1995, 310-311). It is highly likely that relationship subclasses are discipline specific, and while common subclasses will be found to occur in most disciplines, each discipline will also have subclasses of its own.

6. The author acknowledges the role of subject access in the linkage system, but has excluded it from this discussion since it is not part of the descriptive cataloging process.

7. References are also used heavily to link variant forms of names and to provide the syndetic structure of subject access; however, neither of these uses are considered bibliographic relationships because they are not linking to a bibliographic entity. In the first instance they link name entity to name entity and in the second instance they link subject entity to subject entity.

8. For a discussion of the role of references versus the role of added entries see (Svenonius, 1996) and (Lubetzky, 1969, 20-22).

9. In several studies that surveyed catalog users about the use of data elements or ways to improve the catalog’s ability to meet their information needs, contents information was cited as an important element. See, for example, Nitecki, 1993; Maltby and Duxbury, 1972; and Palmer, 1972, 88-92.

10. Although MARBI has approved separate subfield codes for names, titles, and other types of information within the 505 contents note field, authority control will not be imposed on any of these subfields at this time.

11. Many authors have presented an overview of catalog use studies or incorporate details of such studies into literature reviews. Those referred to by this author include Palmer, 1972; Weintraub, 1979; Simpson, 1989; Hufford, 1991; Yee, 1991; Seymour, 1991; Barry, 1994; and Hert, 1996.

12. See, for example, Hufford, 1991; Ballard, 1994; Connaway, Budd, and Kochtanek, 1995; Wyly, 1996; and Lundgren and Simpson, 1997.

13. See, for example, Barry, 1994; Nitecki, 1993; Michel, 1994; Kuhlthau, 1993; and Hert, 1996, among others.

14. For a taxonomy of name relationships see Tillett, 1989b, 10-11.

15. The lack of authority records found in the LC Authority File was in large part due to the Library of Congress' past policy not to create an authority record for works when no references were required, even though a uniform title may be used. This presented special problems when cataloging music with non-distinctive form titles (e.g., quartets, symphonies, etc.), for each cataloger had to construct the multi-element uniform title from scratch for such works, a time consuming and often redundant task that could be eliminated by the creation of a work authority record. LC has recently changed this policy for music materials and will provide such uniform title authority records in the future (as per communication with Barbara Tillett, July 17, 1997).

16. The MCD reads as follows: "Generally the heading referred to should include only the basic uniform title of the work, without additions such as 'arr.' (25.35C), 'Vocal score' (25.35D), 'Libretto' (25.35E), language (25.35F), etc., even if such additions are used in the uniform title in the bibliographic record for the item being cataloged. If, however, the title being referred from is specific to the arrangement, format, language, etc. brought out by an addition to the uniform title, and the title would not logically be used for a different manifestation of the work, refer to the uniform title with the addition." (*Music Cataloging Bulletin* (1990). Vol. 21, no. 6: 4).

17. One of the earliest suggestions for the use of a relational database design for library catalogs came from Gorman (1978 and 1979). Since then other researchers have examined the relational database concept in various contexts including Crawford, Becker and Ogilvie (1984), Chang (1992), Llorens and Trénor (1993), and Green (1996), to name a few.

18. See pages 6-7 of this paper for a list of the functional requirements identified by the seminar group.

19. For example, the Europagate Project has developed gateway software that implements multiple target catalog searches using a World-Wide-Web front end interface. Project information is available at: [URL: <http://europagate.dtv.dk/>].

20. Information on the current Bradford OPAC 2 is available at:
[URL: <http://www.comp.brad.ac.uk/research/database/bopac2.html>].