Collection/Item Metadata Relationships

Allen H. Renear
Graduate School of Library and Information Science
Center for Informatics Research in Science and Scholarship
University of Illinois, USA
renear@illinois.edu

Karen M. Wickett
Graduate School of Library and Information Science
Center for Informatics Research in Science and Scholarship
University of Illinois, USA
wickett2@illinois.edu

Richard J. Urban
Graduate School of Library and Information Science
Center for Informatics Research in Science and Scholarship
University of Illinois, USA
rjurban@illinois.edu

David Dubin
Graduate School of Library and Information Science
Center for Informatics Research in Science and Scholarship
University of Illinois, USA
ddubin@illinois.edu

Sarah L. Shreeves
Graduate School of Library and Information Science
Center for Informatics Research in Science and Scholarship
University of Illinois, USA
sshreeve@illinois.edu

Abstract

Contemporary retrieval systems, which search across collections, usually ignore collection-level metadata. Alternative approaches, exploiting collection-level information, will require an understanding of the various kinds of relationships that can obtain between collection-level and item-level metadata. This paper outlines the problem and describes a project that is developing a logic-based framework for classifying collection/item metadata relationships. This framework will support (i) metadata specification developers defining metadata elements, (ii) metadata creators describing objects, and (iii) system designers implementing systems that take advantage of collection-level metadata. We present three examples of collection/item metadata relationship categories, *attribute/value-propagation*, *value-propagation*, and *value-constraint* and show that even in these simple cases a precise formulation requires modal notions in addition to first-order logic. These formulations are related to recent work in information retrieval and ontology evaluation.

Keywords: metadata; Dublin Core; collections; context; logic; inferencing

1. Introduction

Collections of texts, images, artifacts, and other cultural objects are often designed to support specific research and scholarly activities. Toward that end collections themselves are carefully developed and described. These collection descriptions indicate such things as the purpose of the collection, its subject, the method of selection, size, nature of contents, coverage, completeness, representativeness, and a wide range of summary characteristics, such as statistical features. This information enables collections to function not just as aggregates of individual data items but as independent entities that are in some sense more than the sum of their parts, as intended by their creators and curators (Curral, Moss & Stuart, 2005; Heaney, 2000; Lagoze, et al. 2006 Lee, 2000, 2005; Palmer, 2004, 2006). Collection-level metadata, which represents this information in computer processable form, is thus critical to the distinctive intellectual and cultural role of collections as something more than a set of individual objects.

Unfortunately, collection-level metadata is often unavailable or ignored by contemporary retrieval and browsing systems, with a corresponding loss in the ability of users to find, understand, and use items in collections (Foulonneau, et al., 2005; Wendler, 2004). Preventing this loss of information is particularly difficult, and particularly important, for "metasearch", where item-level descriptions are retrieved from a number of different collections simultaneously, as is the case in the increasingly distributed search environment of the Internet (Christenson & Tennant, 2005; Dempsey, 2005; DLF, 2005; Foulonneau, et al., 2005; Lagoze, et al., 2006; Warner, et al., 2007).

The now familiar example of this challenge is the "on a horse" problem, where a collection with the collection-level subject "Theodore Roosevelt" has a photograph with the item-level annotation "on a horse" (Wendler, 2004). Item-level access across multiple collections (as provided not only by popular Internet search engines, but also specialized metasearch and federating systems, such as OAI portals) will not allow the user to effectively use a query with keywords "Roosevelt" and "horse" to find this item, or, if the item is retrieved using item-level metadata alone, to then use collection-level information to identify the person on the horse as Roosevelt.

The problem is more complicated and consequential than the example suggests and the lack of a systematic understanding of the logical relationships between collection-level metadata and item-level metadata is an obstacle to the development of remedies. This understanding is what is required not only to guide the development of context-aware search and exploitation, but to support curation policies as well.

The problem is also urgent: even as recent research confirms the key role that collection context plays in the scholarly use of information resources (Brockman, et al., 2001; Palmer, 2004), the Internet has made the context-free searching of multiple collections routine.

We are developing a framework for classifying and formalizing collection/item metadata relationships and determining inference rules that can be incorporated into retrieval and browsing systems. This undertaking is part of a larger project, recently funded by U.S. Institute for Museum and Library Services (IMLS), to develop tools for improved retrieval and exploitation across multiple collections.[24]

2. The DCC/CIMR Project

These issues were initially raised during an IMLS Digital Collections and Content (DCC) project, begun at the UIUC in 2003. That project developed a collection-level metadata schema based on the RSLP and Dublin Core Metadata Initiative (DCMI) and created a collection registry for all digital collections funded through the IMLS National Leadership Grant (NLG) since 1998, with some Library Services and Technology Act (LSTA) funded collections included. The registry currently contains records for 200 collections. An item-level metadata repository was also developed, which has harvested 76 collections using the OAI-PMH protocol. Our research initially focused on overcoming the technical challenges of aggregating large heterogeneous collections of item-level records and collection descriptions. We conducted studies on how content contributors conceived of the roles of collection descriptions in digital environments (Palmer & Knutson, 2004; Palmer et al., 2006), and preliminary usability work. These studies and related work on the CIC Metadata Portal[25], suggest that while the boundaries around digital collections are often blurry, many features of collections are important for helping users navigate and exploit large federated repositories, and that collection and item-level descriptions should work in concert to benefit certain kinds of user queries (Foulonneau, et al., 2005).

Concurrently, we studied the quality of the harvested item-level metadata using a range of qualitative and quantitative metrics. While the obstacles to building effective aggregations of item-level metadata are well documented (Arms et al., 2003; Dushay and Hillmann, 2003; Hutt

[24] IMLS Digital Collections and Content. http://imlsdcc.grainger.uiuc.edu/about.asp

[25] http://cicharvest.grainger.uiuc.edu/

and Riley, 2005), we were interested the quality dimensions that could be measured in order to better understand where poor quality might impede interoperability. Using an information quality framework proposed by Gasser and Stvilia (Gasser and Stvilia 2001; Stvilia et al. 2004) we found that the relational or contextual information quality dimensions—that is, the dimensions that depend on relationships between the information and an aspect of its use or context—were particularly problematic (Shreeves et al., 2005). Unlike intrinsic information quality dimensions in which the information can be measured in relation to a reference standard (such as a date encoding standard), measurement of relational quality dimensions are dependent on what context an item was meant for and its use within that context. In this environment, collection-level metadata could supply some of that context, given a better understanding of the relationships between collection and item level metadata.

In 2007 we received a new three year IMLS grant to continue the development of the registry and to explore how a formal description of collection/item metadata relationships could help registry users locate and use digital items. This latter activity, CIMR, (Collection/Item Metadata Relationships), consists of three overlapping phases. The first phase is developing a logic-based framework of collection/item metadata relationships that classifies metadata into categories with associated rules for propagating or constraining information between collection and item levels. Next we will conduct empirical studies to see if our conjectured taxonomy matches the understanding and behavior of metadata creators, metadata specification designers, and registry users. Finally we will design and implement pilot applications using the relationship rules to support searching, browsing, and navigation of the DCC Registry. We will also suggest OWL[26] bindings for the categories and inference rules. Although this framework will be applicable to collection-level descriptions generally, our initial focus is on the Dublin Core Collections Application Profile (DCMI, 2007).

The collection/item metadata relationships framework will allow metadata specification designers to more precisely indicate the relationships intended or assumed by their specifications. These applications of the framework are explicit classifications of metadata elements which will in turn provide guidance both to metadata creators assigning metadata and to systems designers mobilizing collection-level metadata in retrieval and browsing systems. In this way the framework supports:

- *Metadata specification developers defining metadata elements*. Metadata specification developers will be able to use applications of the framework to indicate the semantics of various metadata elements in their specifications.
- *Metadata creators describing objects*. Metadata librarians can use applications of the framework to confirm their understanding of the metadata elements they are assigning.
- *Systems designers developing and configuring retrieval systems*. Software architects can use applications of the framework to guide the design and implementation of automatic inferencing features in retrieval and browsing software.

In addition collection curators can use applications of the framework to improve metadata quality by discovering inconsistencies in metadata assignments between the collection and item levels, and to facilitate semantic interoperability with other databases and applications.

Many benefits of such a framework can be realized almost immediately. Later, when formal specifications and tools based on them are in place, the intended relationships (specified in a computer processable formats) can be integrated directly into management and use, as well as software. However realizing this level of value will require not only completing a plausible framework of relationships, but developing a public specification that is practical and reflects the

[26] http://www.w3.org/TR/owl-features/

common understandings of the metadata community. The current paper is only a first step in that direction. [27]

3. Three Kinds of Metadata Relationships

Currently we are focusing on defining categories for the simplest cases, where information recorded at the collection level can be usefully, if not always completely, converted to information at the item level. So far we have identified three categories, *attribute/value-propagation*, *value-propagation*, and *value-constraint,* which will serve to illustrate our approach.

Our characterizations are being developed in first order logic, extended as necessary by modal notions and other constructs. This is partly to ensure precision and clarity, and partly in anticipation of a final specification in RDF/OWL that will support automatic inferencing. However we work initially in first order logic rather than directly in OWL in order to take advantage of a compact familiar notation with well-understood semantics, and which can be easily extended as necessary to include modal, temporal, or other features. Since the use of first order logic with extensions will allow the expressiveness of our characterizations to be greater than that available in the appropriate level of OWL, a reductive strategy may be in order when we begin those translations.

3.1. Attribute/Value Propagation

Consider the DC Collections AP property *marcrel:OWN*, adapted from the MARC cataloging record standard. It is plausible that within many legal and institutional contexts whoever owns a collection owns each of the items in the collection, and so if a collection has a value for the *marcrel:OWN* attribute then each member of the collection will have the same value for *marcrel:OWN*. (For the purpose of our example it doesn't matter whether or not this is actually true of *marcrel:OWN*, only that some attributes are sometimes used by metadata creators with an understanding of this sort, while others, such as *dc:identifier*, are not). We refer to this meta-property of metadata elements as *attribute/value propagation* (or *a/v-propagation*). An informal definition might be:

Def a/v-p 1: an attribute **A** *a/v-propagates* =df
if a collection has some value *z* for **A**, then each item in the collection has *z* for **A**.

Some collection-level metadata elements a/v-propagate to collection members, and some don't — those that do present obvious opportunities to preserve context by bringing collection-level information to the item level.

A natural formalization of **Def a/v-p 1** in first order logic would be:

Def a/v-p 2: An attribute **A** *a/v-propagates* =df
$\forall x \forall y \forall z$ [(IsGatheredInto(x,y) & A(y,z)) $\supset$ A(x,z)]

Here we use *IsGatheredInto*, from the DCMI Collections AP to represent the item/collection relationship (DCMI, 2007). We assume that if something x IsGatheredInto something y then y is a collection and x is a member (of a collection). Or in the notation of first order logic: $\forall x \forall y$ [IsGatheredInto(x,y) $\supset$ (Member(x) & Collection(y))].

3.2. Interlude I: Propagation vs. Inheritance

Although attribute/value propagation from collection to members might be considered a kind of inheritance, in some very broad sense of inheritance, we think it is misleading to classify it as

[27] A briefer description of CIMR at an earlier stage of development is Renear et al. (2008a).

such. A little analysis shows that attribute/value propagation is in any event clearly not classical subsumptive inheritance as found in frame-based systems and semantic networks.

Consider a typical example of a taxonomic class hierarchy: Fido is an *instance of* the class DOG; DOG *is a subclass* of MAMMAL; and MAMMAL has the attribute/value pair *thermeoregulation=warmblooded*. DOG inherits *thermeoregulation=warmblooded* from MAMMAL in virtue of the fact that DOG is a subclass of (*a kind of*) MAMMAL; and that Fido inherits (although not in precisely the same sense) *thermeoregulation=warmblooded* from DOG because Fido is an instance of (*is a*) DOG. Note that there are two sorts of inheritance supporting relationships in our example: *subclass* and *instance*. The classical notion of inheritance has varying interpretations and ambiguities (Woods, 1975; Brachman, 1983), but in any case it is easy to see that neither of these two inheritance-supporting relationships, subclass and instance, matches the *IsGatheredInto* relationship between items and their collections: a member of a collection is neither a *subclass of* a collection nor an *instance of* that collection.

Our use of the term "propagation" in this sense is intended to follow Brachman (1991).

3.3. Value Propagation

Another collection/item metadata relationship is almost, but not quite, this simple. Consider the collection-level attribute *mycld:itemType*, intended to characterize the type of objects in a collection, with values from the DCMI Type Vocabulary (for the example we assume homogeneous collections, so this is an additional refinement on DCMI *cld:itemType*). Here we cannot conclude that if a collection has the value *dcterms:Image* for *mycld:itemType* then the items in that collection also have the value *dcterms:Image* for that same attribute. This is because an item that is an image is not itself a collection of images and therefore cannot have a value for *mycld:itemType*.

However, while the rule for propagating the information represented by *mycld:itemType* from collections to items is not simple propagation of attribute and value, it is nevertheless simple enough: if a collection has a value, say *dcterms:Image*, for *mycld:itemType*, then the items in the collection have the same value for a corresponding attribute, say, *dc:type*. The metadata elements *mycld:itemType* and *dc:type* have the same domain of values, but a different semantics. When two metadata attributes are related in this way we say the first *value-propagates* (or *v-propagates*) to the second. Informally:

Def v-p 1: an attribute **A** *v-propagates* to an attribute **B** =df
if a collection has the value *z* for **A**, then every item in the collection has the value *z* for **B**.

Notice that in this view, a/v-propagation is a special case of v-propagation: an attribute a/v-propagates precisely when it v-propagates to itself.

A formalization of **Def a/v-p 1** in the symbolism of first order logic would be:

Def v-p 2: An attribute **A** *v-propagates* to an attribute **B** =df
$\forall x \forall y \forall z$ [(IsGatheredInto(x,y) & $\mathbf{A}$(y,z)) $\supset$ $\mathbf{B}$(x,z)]

3.4. Value Constraints

Some collection/item metadata relationships are less direct than simple value propagation. In these cases, the value for the attribute on the item level is not the same, but does stand in some particular relation to the value for the collection-level attribute. For example, consider the collection-level attribute *mycld:dateItemsCreated* from the DC Collections AP, and the item-level attribute *mydc:created*. If a collection has a date range given as the value for *mycld:dateItemsCreated*, then we can infer about each item in that collection that a date given for the value of *mydc:created* will fall within that date range (for this example we assume neither of these attributes may be repeated, so these are again a refinement of the corresponding DCMI terms). We refer to these cases as *value constraints* (or *v-constraints*), since the collection-level metadata can be seen as constraining the values for a particular item-level attribute.

Informally:

Def v-c 1: an attribute **A** *v-constrains* an attribute **B** with respect to a constraint **C** =df if a collection has the value z for **A** and an item in the collection has the value w for **B**, then w is related to z by **C**.

The predicate variable **C** in the definition above represents the constraint between the values and will vary with the semantics of the related attributes. The constraint discussed in the example above is temporal containment, other sorts of constraints would be relevant to other sorts of metadata elements — for instance, spatial metadata might have spatial containment constraints. The modeling of this kind of metadata relationship may be useful for validation of item-level metadata in regard to the intent of the metadata creators.

A natural formalization for v-constraint would be:

Def v-c 2: an attribute **A** *v-constrains* an attribute **B** with respect to a constraint **C** =df

$$\forall x \forall y \forall z \forall w\ [(\text{IsGatheredInto}(x,y)\ \&\ \mathbf{A}(y,z)\ \&\ \mathbf{B}(x,w)) \supset \mathbf{C}(w,z)]$$

3.5. Interlude II: The Need for Modalization

Since the formalizations **Def a/v-p 2, Def v-p 2, and Def v-c 2** use truth-functional material conditionals ("P ⊃ Q") to express the conditional assertions seen in **Def a/v-p 1 Def v-p 1**, and **Def v-c 1** they fell prey to familiar difficulties sometimes referred to as the "paradoxes of material implication." The so-called paradoxes are the counterintuitive results that follow from the truth functional material conditional being defined as true whenever the antecedent is false (regardless of the truth value of the consequent), and whenever the consequent is true (regardless of the truth value of antecedent).

Consider the attribute, *acme:collIdentifier*, whose value is intended to be a collection identifier assigned by a particular identifier assignment agency, the ACME collection identifier agency. This attribute is obviously not a/v-propagating: one cannot conclude from the fact that a collection has a value for *acme:collIdentifier* that the items in the collection have that value (or even any value) for *acme:collIdentifier.* However before the assignment of any of these collection identifiers by the ACME agency there will be no collections with a value for *acme:collIdentifier*. Therefore, the conditional will be satisfied ("trivially") and *acme:collIdentifier* will be classified as a/v-propagating, which it is not.

To avoid this erroneous result, we can use a modal version of the conditional which, in the case of a/v-propagation, states that an attribute **A** a/v-propagates if and only if *it is impossible for*: a collection to have v for **A** and its items not have v for **A**.

Def a/v-p 2: An attribute **A** *a/v-propagates* =df

$$\Box \forall x \forall y \forall z\ [(\text{IsGatheredInto}(x,y)\ \&\ \text{A}(y,z)) \supset \text{A}(x,z)\ ]$$

Where the "□" is read "necessarily…".

However although this definition seems like a natural account of a/v propagation and does address the problem with attributes such as *acme:collIdentifier*, it still does not accurately identify all and only attributes that are (intuitively) a/v propagating. This is because modalized conditionals are themselves susceptible to a modal version of the paradoxes of material implication, sometimes called "the paradoxes of strict implication": if the antecedent of a modal conditional is *necessarily* false, then the conditional is true regardless of the consequent; and if the consequent is *necessarily* true, then the conditional is true, regardless of the antecedent. Our approach to this (also well-known) problem is to use preemptive modal restrictions to exclude the remaining counterexamples. A prose version of such a definition might be

Def a/v-p 4: An attribute **A** *a/v-propagates* =df
I. a) It is possible for a collection to have a value for **A**; &
b) It is possible for a collection member to have a value for **A**; &
c) It is possible that some value for **A** is had by one thing and lacked by another; &
II. Necessarily, if some item is a member of a collection which has some value for **A**, then that item has that value for **A**.

Or, in first order modal logic:

Def a/v-p 4: An attribute **A** *a/v-propagates* =df
I. a) $\Diamond\ \exists y \exists z$ [Collection(y) & $\mathbf{A}(y,z)$] &
b) $\Diamond\ \exists x \exists z$ [Member(x) & ~$\mathbf{A}(x,z)$] &
c) $\Diamond\ \exists x \exists y \exists z$ [$\mathbf{A}(x,z)$ & ~$\mathbf{A}(y,z)$] &
II. $\Box\ \forall x \forall y \forall z$ [(IsGatheredInto(x,y) & $\mathbf{A}(y,z)$) $\supset$ $\mathbf{A}(x,z)$].

Where "$\Diamond$" is read "it is possible that…" and is equivalent to "~$\Box$~", Similar modal definitions can be developed for v-propagates and v-constrains. For the rationale for these additional clauses see Renear et al. (2008b).

The problem of trivial satisfaction has been noted in the information retrieval literature, where van Rijsbergen (1986) and Lalmas (1998) argue that it is serious problem, and Sebastiani (1998) argues that it is not, claiming that the conditionals in question do not nest at the level where problems are created. Our analysis seems to support van Rijsbergen and Lalmas, at least for the applications being considered here. When conditionals are used in definitions, or in specification design and conceptual analysis, they do indeed nest at the problematic level, and in the problematic location (the definiens of a definition, or, more generally, in the antecedent of a larger conditional (when "=df" is read "if and only if").

Our particular solution to the problem, a combination of a modalized conditional and preemptive modal exclusion, suggests that any adequate representation of collection/item relationships will require modal notions. We note that our technique of modal exclusion is similar in some respects to the modal "metaproperty" strategy for ontology design (Guarino & Welty, 2004), where modal notions are also used to capture our intuitive understanding of fundamental concepts. We have discussed this problem in further detail elsewhere (Renear, et al., 2008b).

4. Future Research Directions

4.1. Extending the Framework

A complete framework for collection/item metadata relationships would cover not only the entailments from single assertions about collections to single assertions about items, but many other collection/item relationships.

Obviously one major division of collection/item metadata relationships is between those that support inferences from collection-level attributes to item-level attributes, and those that support inferences from item-level attributes to collection-level attributes. In this paper we have given examples of the former sort of relationship only.

Moreover, so far we have only considered cases where the assertion of a single metadata attribute at one level implied the assertion of a single metadata attribute at the other. But a complete framework for collection/item relationship categories must also accommodate the more general case, where assertions of one or more than one metadata attribute at one level imply assertions of one or more than one metadata attribute at the other level.

4.2. Intentionality

Throughout the discussion above we have carefully avoided directly raising questions such as “what is a collection?” and “what is it for something to be *gathered into* something else?”. This is in part because we believe that answering those questions will necessarily involve the current analysis, and so consequently those questions are not genuinely prior, methodologically speaking, to our analysis of collection/item metadata relationships. In fact we see our analysis of collection/item metadata relationships as itself a substantive contribution to questions such as “what is a collection?”. But in any case we cannot long avoid directly addressing the fundamental issue of the role of curatorial intent, which must be part of any analysis of the concept of a collection. When we do take up these issues directly it is quite likely that we will need to extend our logic further, to include intentional as well as alethic modal operators.

4.3. Reduction-Resistant Collection Level Properties

It would seem that some collection-level properties can be safely re-expressed as item-level metadata without loss of information. For instance, if a collection is described as being a collection of images we can (at least arguably) assume that nothing further is intended by that description than that each item in the collection is an image. In this case a/v-propagation and v-propagation carry all intended collection-level information to the item level and can straightforwardly support enhanced discovery and use.

However other sorts of collection-level information cannot be so easily reassigned to the item level without loss of meaning. In such cases the strategy of moving information from the collection level to the item level may still be valuable, but cannot, by itself, fully exploit the information provided at the collection level. Intriguingly these attributes often turn out to be carrying information that is tightly tied to the distinctive role the collection is intended to play in the support of research and scholarship. Obvious examples are metadata indicating that a collection was developed according to some particular method, designed for some particular purpose, representative in some respect of a domain, has certain summary statistical features, and so on. Such features cannot be converted to facts about individual items, and yet this is precisely the kind of information that makes a collection, as a *collection*, valuable to researchers — and if it is lost or inaccessible the collection cannot be useful in the way originally intended by its creators.

Understanding and exploiting metadata of this kind will be a particular challenge.

Acknowledgements

This research is supported by a 2007 IMLS NLG Research & Demonstration grant hosted by the GSLIS Center for Informatics Research in Science and Scholarship (CIRSS), Carole L. Palmer, Principal Investigator. Project documentation is available at http://imlsdcc.grainger.uiuc.edu/about.asp. We have benefited considerably from discussions with other DCC/CIMR project members and with participants in the IMLS DCC Metadata Roundtable, including: Timothy W. Cole, Thomas Dousa, Myung-Jan Han, Amy Jackson, Mark Newton , Carole L. Palmer, Sara L. Shreeves, Michael Twidale, Oksana Zavalina, and Wu Zhang.

References

Arms, William Y., Naomi Dushay, Dave Fulker, and Carl Lagoze. (2003). A case study in metadata harvesting: The NSDL. *Library Hi Tech, 21*(2), 228–237.

Brachman, Ronald J. (1983). What ISA is and isn't: An analysis of taxonomic links in semantic networks. *IEEE Computer, 16* (10), 30-6.

Brachman, Ronald J., Deborah L. McGuinness, Peter F. Patel-Schneider, Lori A. Resnick, and Alex Borgida. (1991). Living with classic: When and how to use a KL-ONE-like language. In John Sowa, *Principles of semantic networks: Explorations in the representation of knowledge,* (pp. 401-456).

Brockman, William, Laura Neumann, Carole L. Palmer, Tonya.J. Tidline. (2001). *Scholarly work in the humanities and the evolving information environment*. Washington, DC: Digital Library Federation/Council on Library and Information Resources.

Christenson, Heather, and Roy Tennant. (2005). *Integrating information resources: Principles, technologies, and approaches*. California Digial Library. Retrieved from http://www.cdlib.org/.

Currall, James, Michael Moss, and Susan Stuart. (2004). What is a collection?. *Archivaria, 58,* 131-146.

Dempsey, Lorcan. (2005, October 9). From metasearch to distributed information environments. Message posted to http://orweblog.oclc.org/archives/000827.html.

DLF. (2005). The distributed library: OAI for digital library aggregation. *OAI Scholars Advisory Panel, 2005 June 20-21, Washington, DC.* Digital Library Federation.

DCMI. (2007). *Dublin Core Collections Application Profile*. Retrieved April 13, 2008, from http://dublincore.org/.

Dushay, Naomi, and Diane I. Hillmann. (2003). Analyzing metadata for effective use and re–use. *DC–2003: Proceedings of the International DCMI Metadata Conference and Workshop,* (pp. 161-170).

Foulonneau, Muriel, Timothy W. Cole, Thomas G. Habing, and Sarah L. Shreeves. (2005). Using collection descriptions to enhance aggregation of harvested item-level metadata. *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries,* (pp. 32-41). ACM Press.

Gasser, Less, Besiki Stvilia, Michael B. Twidale, and Linda C. Smith. (2001). *A new framework for information quality assessment*. Technical report ISRN UIUCLIS--2001/1+AMAS. Champaign, Ill.: University of Illinois at Urbana Champaign.

Guarino, Nicola, and Christopher A. Welty. (2004). An overview of OntoClean. In Steffen Staab and Rudi Studer, *Handbook on Ontologies*. Springer.

Heaney, Michael. (2000). *An analytic model of collections and their catalogues*. UK Office for Library and Information Science.

Hutt, Arven, and Jenn Riley. (2005). Semantics and syntax of Dublin Core usage in Open Archives Initiative data providers of cultural heritage materials. *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries, Denver, Colo, June 7–June 11,* (pp. 262–270). New York: ACM Press.

Lagoze, Carl, Dean Krafft, Tim Cornwell, Naomi Dushay, Dean Eckstrom, and John Saylor. (2006). Metadata aggregation and "automated digital libraries": A retrospective on the NSDL experience. *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries*. New York: ACM Press.

Lalmas, Mounia. (1998). Logical models in information retrieval. *Information Processing and Management*, (*34)*1, 19-33.

Lee, Hur-Li. (2005). The Concept of Collection from the User's Perspective. *Library Quarterly, 75*(1), 67-85.

Lee, Hur-Li. (2000). What is a collection? *JASIS, 51*(12), 1106-1113.

Palmer, Carole L. (2004). Thematic research collections. In Susan Schreibman, Raymond G. Siemens, and John Unsworth (Eds.), *Companion to digital humanities* (pp. 348-365). Oxford: Blackwell Publishing.

Palmer, Carole L., and Ellen M. Knutson. (2004). Metadata practices and implications for federated collections. *Proceedings of the 67th ASIS&T Annual Meeting*.

Palmer, Carole L., Ellen M. Knutson, Michael Twidale, and Oksana Zavalina. (2006). Collection definition in federated digital resource development. *Proceedings of the 69th ASIS&T Annual Meeting, Austin, Texas, 2006.*

Renear, Allen. H., Richard Urban, Karen Wickett, C. L. Palmer., and David Dubin. (2008a). Substaining collection value: Managing collection/item metadata relationships. *Proceedings of the Digital Humanities conference, Oulu, Finland, 2008.*

Renear, Allen. H., Richard Urban, Karen Wickett, and David Dubin. (2008b). The return of the trivial: Formalizing collection/item metadata relationships. *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries 2008*. New York: ACM Press.

Sebastiani, Fabrizio. (1998). On the Role of Logic in Information Retrieval. *Information Processing and Management 34*(1), 1-18.

Shreeves, Sarah L. , Ellen M. Knutson, Besiki Stilva, Carole L. Palmer, Michael B. Twidale, and Timothy W. Cole. (2005). Is 'Quality' Metadata, 'Shareable' Metadata? The Implications of local metadata practices for federated collections. *Proceedings of the Twelfth National Conference of the Association of College and Research Libraries, Minneapolis, 2005, (*pp. 223-237). Chicago, IL: Association of College and Research Libraries.

Stvilia, Besiki, Less Gasser, Michael B. Twidale, Sarah L. Shreeves, and Thomas W. Cole. (2004). Metadata quality for federated collections. *Proceedings of ICIQ 04-9th International Conference on Information Quality*, *Cambridge, MA*.

Van Rijsbergen, C. J. (1986). A non-classical logic for information retrieval. *The Computer Journal 29*(6), 481-485.

Warner, Simon., Jeroen Bekaert, Carl Lagoze, Xiaoming Lin, Sandy Payette, and Herbert Van de Sompel. (2007). Pathways: Augmenting interoperability across scholarly repositories. *International Journal on Digital Libraries*. doi: 10.1007/s00799-007-0016-7

Wendler, Robin. (2004). The eye of the beholder: Challenges of image description and access at Harvard. In Diane I. Hillmann, and Elaine L. Westbrooks (eds.), *Metadata in Practice,* (pp. 51-6). Chicago: American Library Association.

Woods, W. A. (1975). What's in a link: Foundations for semantic networks. In Daniel G. Bobrow, and Allan Collins (eds.), *Representation and Understanding: Studies in cognitive science*. New York: Academic Press.