Interrogating "Identity": A Philosophical Approach to an Enduring Issue in Knowledge Organization */**

Jonathan Furner

University of California, Los Angeles, UCLA-GSE&IS, 300 Young Dr N, Mailbox 951520, Los Angeles, CA 90095-1520, USA <furner@gseis.ucla.edu>

Jonathan Furner is an associate professor in the Graduate School of Education and Information Studies at the University of California, Los Angeles. He has a Ph.D. in information studies from the University of Sheffield, and an M.A. in philosophy and social theory from the University of Cambridge. He works on cultural informatics, and the history and philosophy of documentation. His current research includes studies of social tagging, art museum documentation, and the metaphysics of documents and their subjects.

Furner, Jonathan. **Interrogating "Identity": A Philosophical Approach to an Enduring Issue in Knowledge Organization.** *Knowledge Organization, 36(1)*, 3-16. 28 references.

ABSTRACT: Empirical evaluation of knowledge organization (KO) systems, and of the tools and techniques that are used to build systems, is a key component of the system design process: our success in building better systems depends at least partly on our ability to measure the goodness of current systems, and to recognize the factors that affect system performance. The basic evaluative question might be expressed quite simply: How good are the representations or models—models of the world, of our knowledge of the world, and/or of expressions of our knowledge of the world—that are produced by our usage of particular KO methods? The straightforwardness of this question is offset by a preliminary need to address metaphysical issues of various kinds, consideration of which can lead us into a quagmire of methodological, epistemological, and ethical problems. What, in this context, is "goodness"? What is the fundamental nature of the kinds of things to be represented? What are the conditions that must be satisfied for a single individual thing to retain its identity over time, and for two individual things to be instances of "the same" kind of thing? Where are the boundaries to be drawn between one thing (or kind of thing) and another? Where does one thing (or kind of thing) stop and another start? How can we come to know the answers to questions about identity, and how we can know when we know? How have we answered questions about identity in different ways at different times and in different places? How ought we to answer questions about identity, and what justifications can we provide in support of our normative claims? As is indicated by the conference organizers' choice of theme for ISKO 2008, designers and evaluators of KO schemes contend on an ongoing basis with issues relating to identity, and a philosophically-informed engagement with such issues is an essential preliminary to understanding evaluation criteria for KO activity. In this talk, the utility for KO of philosophical theories of identity is examined, and motivation is provided for the additional use of such a philosophical framework in evaluating the extent to which KO schemes successfully reflect the cultural identities of their users.

* Keynote speech delivered at the 10th International ISKO Conference (Montréal, Canada, August 5-8, 2008) August 6, 2008.

** **Acknowledgements:** I would like to thank Clément Arsenault and Joe Tennis, conference chair and program chair respectively, for the very kind invitation that resulted in my giving this talk at the 10th International ISKO Conference.

1.0 Introduction

The topic of this paper is the significance for knowledge organization of analyses of the concept of identity. My motivation for choosing this topic is the theme of the 10th International ISKO Conference, which is "Culture and identity in knowledge organization." One claim that is implicit in that choice of theme is that there exists at least one issue that has to do with identity and that has an impact on the goals of researchers in knowledge organization (KO) or the results of their work. I am going to take the liberty in

this paper of suggesting one way in which that claim may be interpreted. My position is that the primary issue is one of *evaluation*. There is a strong claim being made here that you may or may not agree with, and I will do a little more to argue for it later. This strong claim is that, ultimately, it is our responsibility as KO researchers to figure out how to build KO systems that work *well*. And to figure out how best to go about *determining* how well KO systems work.

In this context, the issue is specifically one of evaluating how KO systems do at handling *identity*. Now, as we shall see, there are many ways in which that word "identity" may be understood, and part of the challenge is untangling all those different senses in which the word may be used. But there are two things we can do immediately to clarify certain aspects of the issue. One of those is to specify that when we talk about how well KO systems deal with identity, we are really talking about one in particular of the functions of KO systems, which is that of *representation*. KO schemes are representations or models of reality. We may disagree vehemently about what things, or what kinds of things, count as *real* things; but I would be surprised to encounter a conception of KO schemes that does not include the representational function as a necessary condition for being a KO scheme. So, the question becomes: How well do KO systems *represent* identity? Secondly, there seems to be an important sense in which we can distinguish between identity singular and identities plural. On the face of it, it seems as if there is a distinction to be made between identity as a relation between things, in the sense in which something might be said to be identical with or *the same* as something, and identities as properties of things, in the sense in which something might be said to *have* a particular identity.

There appears to be a way of breaking down the basic question—How well do KO systems represent identity?—into two separate questions that correspond respectively to senses of identity as relation and identity as property. In the first place, we might ask, How well do KO systems represent relationships of identity between classes of documents? And how well do KO systems help indexers *and* searchers explore those relationships? In the second place, we might ask, How well do KO systems help indexers, classifiers, catalogers organize knowledge about the personal or social identities of members of social groups? How well do KO systems help people find the right labels for classes of documents that are about those identities, and help people find those documents? Even though both of these kinds of question are about identity, they are often treated quite separately in the KO literature. Sometimes the relationship between them is emphasized, but certainly not always. One of the objectives of this paper is to demonstrate that it is at least somewhat helpful to emphasize the relationship by considering the two questions in tandem. And a subsidiary objective is to show that it is at least possible, if not desirable, to do this using a conceptual framework that looks to philosophy on the one hand, and information retrieval on the other, for inspiration.

The primary objective, however, is to defend a series of related propositions. These are as follows:

- firstly, that identity is analyzable in a way that can inform our decisions about how to analyze two other relations that have historically been considered very important in KO, and they are *aboutness* and *relevance*;
- secondly, that the *production* of identity, in a sense that will be explained later, could usefully be considered to be the ultimate *goal* of KO;
- thirdly, that achieving the effective representation in KO systems of *personal and social identities* is a complex special case of a general challenge facing some traditional KO techniques; and thus
- that the concept of identity is central to KO, possibly even more central than its selection as this particular conference's theme indicates.

The approach that I will be taking is one that involves analysis of the *concepts* that we use to talk about issues. In my characterization of the main issue as one of determining how well KO systems do at representing identity, there are four core concepts: (i) identity; (ii) knowledge organization; (iii) representation; and (iv) goodness. The sense of goodness that I use here is just the sense in which some KO systems are good, some are bad, and therefore some are better than others, at doing certain things. This is very similar to the sense in which María López-Huertas uses "quality" in her contribution to the special issue of *Knowledge Organization* on the topic "What is knowledge organization?" (López-Huertas 2008). María López-Huertas explicitly identifies "quality" as something to aim for in the design of KO systems, and the implication is that evaluation of quality is an absolutely necessary component of the KO system design process.

2.0 Identity

So, first, to *identity* itself. The magnitude of the challenge here might be demonstrated simply by listing some of the kinds of identity that are dealt with in the various literatures (see Table 1). I did make some effort to be exhaustive with this listing, but I still have the feeling that this is really just the tip of the iceberg. In the first place we have various conceptions of individual, personal, or self identity, to be distinguished from various conceptions of group, collective, or social identity. We can distinguish the different kinds of identity defined in different domains of theory and practice, such as cultural, economic, and psychological identity. If we focus on personal and social identity, there are all sorts of dimensions on which different identities may be distinguished, so that we talk about racial, sexual, and linguistic identity, and so on. If we take a metaphysical or logical approach to identity, we find that it is possible to distinguish numerical and qualitative, relative and absolute, synchronic and diachronic identity.

Individual, personal, self
Group, collective, shared, communal, community, social
Cultural, political, economic, psychological, legal, metaphysical, logical, mathematical
Racial, ethnic, gender, sexual, national, linguistic, religious, professional, occupational, familial
Numerical, qualitative, relative, absolute, synchronic, diachronic, transworld
Disciplinary, institutional, departmental
Corporate, brand, product, visual
Mistaken, split
Digital, electronic, virtual, online
User, object, work, bibliographic, record, citation

Table 1. Kinds of identity.

In addition, there are a whole slew of related concepts (see Table 2). These are not *kinds* of identity, so much as concepts whose meaning could usefully be clarified in any analysis of identity. There are concepts that seem to have very similar meanings to identity, such as sameness, identicality, similarity, and indiscernibility. There are concepts that seem to have opposite meanings, such as individuality, uniqueness, distinctness, difference, and diversity. And then there are lots of things that can be done to and with and by and through identity and identities—including organization, classification, and categorization, of course. A conceptual minefield!

User identifier, object identifier, work identifier, record identifier
Identity problem, theory, politics, crisis, theft, status, card
Sense of identity
Property, relation, image, role
Privacy, security, confidentiality, trust, reputation, verification, authentication
Sameness, identicality, similarity, indiscernibility
Individuality, uniqueness, distinctness, difference, diversity
Authenticity, cohesion, coherence, tolerance, hybridity
Formation, construction, capture, representation, exploration, manipulation, management
Identification, individuation, differentiation, discrimination, instantiation, exemplification, characterization
Organization, classification, categorization

Table 2. Related concepts.

Even if we limit ourselves to looking at philosophical approaches to the study of identity (see, e.g., Noonan 2006; Heyes 2007), the literature is enormous and varied and scattered very widely. Different analytical approaches have been taken in philosophy of logic, metaphysics, social and political philosophy, philosophy of technology, and philosophy of art, as well as in what we might call philosophy of documentation or even philosophy of knowledge organization.

Let us return to the basic distinction made earlier, between identity conceived as a relation and identity conceived as a property. Sometimes we talk about "the identity of *x* and *y*"; sometimes we talk about "the identity of *x*." What is going on here?

2.1 Identity as a relation

Taking identity as a relation first, here are definitions of two senses of identity. First of all, we say that object *x* and object *y* are *numerically* identical if *x* is the same object as *y*. Notice that it does sound a little odd if we say "two objects, *x* and *y*, are identical." It sounds odd precisely because, in the case of identity, we do not have two objects. The whole point is, we just have one. In fact, more generally, we might say that *x* and *y* are identical if they are countable as one thing. This is why identity in this sense is sometimes known specifically as numerical identity. Secondly, we can contrast numerical identity with qualitative indiscernibility. We say that *x* and *y* are qualitatively indiscernible if *x* has all and only the same properties as *y*. And here is some more terminology: If *x* is *not* the same object as *y*, then we say that *x* and *y* are numeri-

cally *distinct* or *individual*. And if *x* does *not* have *all and only* the same properties as *y*, then we say that *x* and *y* are qualitatively *discernible* or *dissimilar*.

So we have introduced two kinds of relation here: continuous relations and binary relations. A relation is a *continuous* relation (i.e., a relation of degree) if its value can be represented by any point on a line. A relation is a *binary* relation if its value can be represented only by one or other of the two poles of a line. Continuous relations are ones like indiscernibility and similarity, where we can happily talk about indiscernibility as a matter of degree. Two things can be more or less indiscernible, more or less similar. In contrast, binary relations are ones like numerical identity. Two things are either identical (in which case they are actually one thing) or they are not.

In his *Discourse on metaphysics* of 1686, the rationalist philosopher Gottfried Leibniz said: "No two substances resemble each other entirely and differ in number alone." From this statement, which has come to be known as Leibniz's Law, modern metaphysicians have derived two principles or theories—the *principle of the ident(ical)ity of indiscernibles*, and the *principle of the indiscernibility of identicals* (see, e.g., Forrest 2006). The principle of the identity of indiscernibles states that, if *x* and *y* are qualitatively indiscernible, then they are numerically identical. This statement is logically equivalent to the statement that only if *x* and *y* are identical are they indiscernible. In other words, indiscernibility is a sufficient condition for identicality, and identicality is a necessary condition for indiscernibility. Correspondingly, the principle of the indiscernibility of identicals is that, if *x* and *y* are numerically identical, then they are qualitatively indiscernible. Taken together, the two principles imply that *x* and *y* are identical *if and only if* they share all and only the same properties.

Even though these principles may appear, on face value, to be reasonable—tautologous even—each of them has actually turned out to be fairly controversial. Some people argue that it *is* possible for two things to resemble each other entirely, thus denying the principle of the identity of indiscernibles; and many people argue that it is possible for *x* and *y* to be numerically identical but qualitatively discernible—for example, when *x* is me-before-this-talk, and *y* is me-after-it. Whether we are convinced by these arguments may depend on what we count as properties. Weak versions of the principles count so-called extrinsic properties (i.e., relations to other objects) among the properties that must be considered when evaluating indiscernibility; strong versions do not count extrinsic properties.

The problem of identity over time—diachronic identity—remains a live issue in philosophical debate (see, e.g., Gallois 2005). The paradox of the ship of Theseus might be familiar in this context. Every day that Theseus's ship is in the harbor, a single plank gets replaced, until after a few years the ship is completely rebuilt: not a single original plank remains. Is it still the ship of Theseus? And suppose, meanwhile, the shipbuilders have been building a new ship out of the replaced planks? Is *that* the ship of Theseus?

This is essential background material for any discussion about *identity conditions* (a.k.a. identity criteria). The questions here are about two kinds of conditions, in fact: about the conditions under which *x* should be considered the *very same thing* as *y*, and about the conditions under which *x* should be considered an instance of the same *kind* of thing that *y* is an instance of. We might like to distinguish on this basis between criteria for *individuation* and criteria for *instantiation*. Although this terminology is not really standardized, it is clear that these are very different questions. And in fact they are questions that lots of people in KO and in the information sciences more generally are interested in, because they are exactly the kinds of questions that need to be answered if we are going to do a good job of designing systems that can determine mechanically whether one document is an instance of the same work, or class, or kind, or type, as another document.

Allen Renear and Richard Smiraglia and others are authorities in this area. Some of Allen Renear's work aims to establish identity conditions for digital objects and clarify, for instance, what it means to say that one version of an electronic document, such as its XML code, is an instance of "the same" document as another version, such as its rendering in a browser (see, e.g., Renear 2007). This work has ramifications for Richard Smiraglia's studies of workhood, and what a FRBRized conception—of the relations between works, expressions, manifestations, and items—means for attempts to specify identity conditions for works (see, e.g., Smiraglia 2001). Here the main question is, Given two items, how can we tell whether they are instances of the same work or of different works? There is a school of thought that these kinds of problems are less problems of identity, as such, as they are problems of workhood, since different kinds of things are going to have different identity conditions, and the interesting thing about works is not *that* they have properties that serve as identity conditions, but *what* those properties are that literally define what it means to be a work.

2.2 Personal and social identity

Now, just as there are many different philosophical, rationalist approaches to the a priori study of identity, of course there are many psycho-socio-cultural approaches to the empirical study of identity, too. People working in developmental psychology, social psychology, cultural anthropology, cultural studies, and political science, as well as in social and political philosophy, have much to say about various aspects of *personal and social identity*, and these are themselves based on some fundamental assumptions about the metaphysical nature of identity, which actually turn out to be quite different from the ones we have just looked at (see, e.g., Abrams 2001; Bilgrami 2001; Ceruso 1997; Hardin 2001; Marcia 2001). It would be worth examining these even if it were just for the purposes of comparison.

To begin, here is a proposal of a primitive definition of personal identity. The personal identity of a person is the property or set of properties that a person has, that *individuates* that person, that distinguishes that person from another—or at least that *instantiates* that person, that identifies that person as an instance of the same kind of person that other persons are instances of, or as a member of the same social group that other persons are members of. Similarly, the identity of a social group is the property or set of properties that a social group has, that identifies that group, that distinguishes that group from another.

There are many different kinds, or facets, of personal and social identity. A non-exhaustive list of some that are considered important by different agents engaged in the process of identification at different times is presented in Table 3. I am sure we could come up with many more. In the figure, they are listed in alphabetical order—because who is to say which is more important than another?

In this discourse, we have started to look at identity more as a property than as a relation. And there is an interesting thing that happens when we start to look at identity as a property rather than as a relation. There is a kind of inversion whereby identity is equated not with sameness, but with the opposite of sameness—that is, with *difference*. Here, the idea is of identity as the property or set of properties that *x* has, in virtue of which it is different and thus distinguishable from *y*. In other words, the identity of *x* is whatever property that *x* has that makes *x* individual, that identifies it.

Age
Ancestral territory
Ancestry/genealogy
Class
Community
Culture
Discipline/field
Ethnicity
Family
Gender
Group
History
Hobby/interest
Home/birthplace
Language
Mental ability
Mythical origin
Nationality
Organization/department
Physical ability
Political party
Profession/occupation
Race/phenotype
Religion
Sexual orientation
Skin color
Society
Subculture

Table 3. Kinds/facets of personal/social identity.

This basic idea of identity, as what makes something individual, seems quite simple. But it turns out we can make it quite complex if we want to. For example, we might have good reason for wanting to conceive the identity of *x* as whatever it is that person *a thinks* is the property that makes *x* individual, or indeed whatever it is that person *b* (who may or may not be the same as person *a*) projects to others as an image of the property that person *a* thinks is the property that makes *x* individual, or whatever it is that person *c* (who may or may not be the same as person *b*) thinks is the image projected by person *b* of the property that person *a* thinks is the property that makes *x* individual ...

One thing that all this highlights is that the process of identification is an active process that is always the result of human intentionality and subjectivity. It is an action carried out by an agent on an object (where object here just means the object of the act), the result of which is the naming of the property that identifies the object (i.e., the class instantiated by the object).

And sometimes—very frequently, in fact—the agent is the same person as the object. In which case, identification is a process of self-categorization, or affiliation with a particular group or category. It might be helpful to assume that, in this process, an agent acts more or less autonomously, that is, free of any *logical* constraints, in the sense that an individual is essentially free to choose whatever identity he or she wants. From a human rights perspective, it could be argued that to be in that position of choice, to define one's identity in whatever way one wishes so long as the rights of others are not thereby infringed, is a basic human right.

The complexities do not end there. It seems as if there are at least eight dimensions on which we can simultaneously locate any given act of identification.

1. In the first place, there is the degree of *subjectivity* assumed to be involved in any act of identification. For example, let us say the object of identification is me. We might allow that it is possible to talk meaningfully about what the identity of me actually, objectively is—about how I really am. We might deny that that is possible, even if we accept that there is a way in which I really am, because it is not clear how any one of us might be able to know how things really are. We might be most interested in talking about what most people think is the case—about the intersubjective consensus. But then again we might be interested in talking about what one particular person (e.g., me) thinks is the case, or what that particular person thinks most people think is the case, or even what that particular person thinks another individual thinks is the case.
2. *Singularity* (of the object). We can distinguish between an act of identification whose object is a single thing, and one whose object is a collection or group made up of multiple things.
3. *Intentionality* (of the object). We can distinguish between an act of identification whose object is a person or group of persons, and one whose object is not a person or group of persons.
4. *Singularity* (of the agent/subject). We can distinguish between an act of identification whose subject—the agent doing the identifying—is a single thing, and one whose subject is a collection or group made up of multiple things.
5. We can look at the *relation* of the subject to the object. Is the agent who is doing the identifying identifying their self, or something other?
6. We can look at the *power* of the agent who is doing the identifying, relative to that of the object. Is the subject in a position of domination over the object, or in a subordinate position?
7. *Particularity*. An act of identification could be one in which one particular thing is (numerically) distinguished from another particular thing, or it could be one in which one kind of thing is (qualitatively) distinguished from another kind of thing.
8. *Intrinsicality* of the identifying property. An act of identification could be one in which the identifying property is considered to be intrinsic to the object, or one in which the identifying property is considered to be extrinsic.

Different fields have different kinds of interest in personal and social identity, but it is possible to identify some general categories of empirical research questions that are consistently attract cross-disciplinary attention: 1. What kinds of *processes*—e.g., affective, behavioral, cognitive—are involved in individuals' affiliating with and prioritizing particular identities? 2. What kinds of *factors* affect individuals' affiliating with and prioritizing particular identities? 3. In what ways and to what extents do individuals' affiliations with and prioritizations of particular identities *affect* the other kinds of decisions and actions taken by those individuals? A large part of this third question relates to the kinds of decisions and actions that are taken that result in the representation, expression, or reflection of individuals' identity affiliations in symbolic form, in documents. The challenge for KO is how to make sure that such expressions of identities are represented in KO systems in ways that serve the users of those systems.

3.0 Knowledge organization, representation, and goodness

Now I would like to say some more about the other three of the four core concepts that I mentioned earlier. For a definition of knowledge organization (KO), we need look no further than Joe Tennis, who says that "KO ... is the field of scholarship concerned with the design, study, and critique of the processes of organizing and representing documents that societies see as worthy of preserving" (Tennis 2008, 103).

Here is another, slightly different take on a definition of KO: KO is the practice and the theory of building KO systems that work well. The intention here is to make a point that is possibly a little more controversial than it might at first seem, which is that

KO *systems* do not *support* KO: they are the *product* of KO. (It may be that here we have two rival conceptions of KO: one that sees KO as a set of processes—organization, representation, and so on—that are ends in themselves, and another that sees KO as the means to a variety of ends, including not only retrieval and access and preservation, but also learning and understanding and mapping and modeling. In practice, of course, the distinction between these two conceptions is fuzzy. Perhaps a more realistic picture would be one of a continuum of conceptions ranging between the two poles.)

There are two central research questions in KO conceived in this way. The first is the *design* question: How ought subjects, and the relations between them, to be represented in a KO system? A standard sort of response to the design question is that subjects and relations should be represented in whatever ways that evaluations tell us are the best. But then there is the *evaluation* question, which is, How do we evaluate? *How do we decide* how subjects, and the relations between them, ought to be represented in a KO system? Let us be clear about it—*this* is the big one. Answer this one, and everything follows.

I think it is possible to distinguish two conceptions of the goal of the practice of KO, and this distinction corresponds roughly to the one Raya Fidel draws between two conceptions of the goal of indexing (Fidel 1994). On the one hand, we have the *document-centered* view that indexers should aim to assign index terms to documents (or documents to index terms) in whichever way it is that produces the most accurate representation of that content. On the other hand, there is the *user-centered* view that indexers should aim to associate documents with those terms that are most likely to be used by searchers looking for those documents.

3.1 Description-oriented KO

Similarly, I think that, on the one hand, we have a *description-oriented* conception of the goal of KO, being to build systems that do well at helping people produce accurate descriptions and representations of documents. And on the other hand, we have a *retrieval-oriented* conception of the goal of KO, being to build systems that do well at helping people find the documents they think they want to find.

Looking more closely at the description-oriented conception first, we might see that it assumes that a necessary condition for a system doing well at helping indexers and classifiers produce accurate representations of documents is that the system itself incorporates an accurate representation of the universe of knowledge that can be used by indexers. We might choose to evaluate that representation or model by assessing how it stacks up against *internal* criteria like coherence, richness, simplicity, or elegance. But, more likely, we would choose to evaluate it by comparing it with *external* criteria—e.g., the way things really are, or the way somebody thinks things are—and seeing what the degree of correspondence or match is. That is how we would decide how "accurate" a representation is.

Let us just take a quick further look at that last criterion: "Correspondence with the way someone thinks things are." We need this criterion as an alternative to "Correspondence with the way things really are" because, even if there were a way in which things really are, nobody, not one of us, could have knowledge of it. This, of course, is an epistemological argument that has excited many people over the years. For now, I will simply assert that different people see reality in different ways, and draw from that the conclusion that every KO system is necessarily and unavoidably "biased," in the sense that every KO system reflects the view of reality of its designers.

To say the least, this is a bit of a problem. We seem to be saying that it is impossible for any KO system either to simultaneously reflect the views of everybody, or even to simultaneously reflect the views of all of its users—which is unfortunate, because of course we would rather like it *not* to be impossible. In fact, many of us would argue that, even if—actually, *especially* if—we accept this description-oriented conception of the goal of KO, then there's another very significant external criterion that should be brought into play when evaluating how well any KO system meets that goal. And that criterion is *justice* or fairness.

There are a number of rival conceptions of social justice—of what it means for any policy or action that affects the multiple members of any group to be evaluated as "just." One category of conceptions is called communitarian (as opposed to contractarian; see Furner 2008). If we apply a communitarian conception of social justice as a criterion for evaluating KO systems, then this is how it plays out. We can say that *the just KO system* is one that supports the distribution of cultural resources without violating the rights or liberties of particular groups or communities and their members—especially minorities and other groups that have historically been oppressed by the dominant groups in power (cf. Beghtol 2005). In other words, the just KO system is one that supports

equitable access to an exhaustively comprehensive range of documents for all members of society, no matter what their motivations are (unless those motivations infringe upon the rights of others).

A just KO system is what we wish for. Unfortunately, just because we know we want it, does not mean that it is possible to get it.

3.2 Retrieval-oriented KO

Let us have a look at the retrieval-oriented conception of the goal of KO. Here the goal is conceived as one of helping users of KO systems—both indexers and searchers—improve the quality of their and other users' access to documents, and benefit from that access. In general, the criteria that we might use to evaluate KO practice might at first sight *look* quite different under this conception than the ones we looked at a couple of slides back. This time it seem like we are not so concerned about correspondence with reality or internal coherence. We are more concerned with the degree to which KO practice produces KO systems that enable access to documents in an *effective*, *efficient*, and *easy* manner. These are criteria that have been examined every which way, in tests of information retrieval (IR) systems.

Effectiveness is usually highlighted as the most important of these retrieval-oriented criteria, and it is worth taking a closer look at the factors that have been identified in IR tests as the ones that have the greatest influence on levels of effectiveness. There are two conceptions of the priority of these factors: an objectivist conception and a user-oriented conception.

In the *objectivist* conception, the factor that has the greatest influence on levels of retrieval effectiveness is the degree of correspondence between the model of reality constructed by the KO system designer, and reality itself. This is interesting. Even though we are adopting a retrieval-oriented conception of the goal of KO at this point—one that suggests that the best way to evaluate KO practice is to determine how well it produces KO systems that enable effective retrieval—we are now saying that the best way of ensuring effective retrieval is to make sure that the model of reality enshrined in the KO system is an accurate representation of reality. In effect, we are recommending using the same criterion to evaluate KO as we did under the description-oriented conception of KO.

In the *user-oriented* conception of the priority of factors affecting retrieval effectiveness, the key factor is the degree of correspondence between the model of reality of constructed by the KO system designer, and the mental model of the world that the KO user has, whether indexer or searcher. The idea is that, if there is any sort of mismatch between the way the world is represented by the KO system and the way the user expects it to be represented, then there is a problem. It does not make any difference how the world actually is: KO system design is all about matching people's images of that world.

This is an idea that is very similar, of course, to the claim that indexing should only be evaluated on the basis of the extent to which the terms assigned to documents by indexers match those that are used by searchers for whom those documents are relevant. It does not make any difference what the documents are actually about—there is not really a way in which it makes sense to say that a document is about anything in particular anyway—all that matters is that indexers are able to predict how searchers will describe documents.

4.0 Aboutness, work-instantiation, and relevance

We saw earlier that there are two conceptions of identity: one of identity as a relation between things, and another of identities as properties of things. We now need to talk about relations in general. Rebecca Green makes it clear: "Relationships are at the very heart of knowledge organization" (Green 2008, 150). There are two ways of distinguishing relations: either we can identify distinctive properties of the *entities* being related, or we can identify distinctive properties that are intrinsic to the *relations* themselves.

In the quaintly-named bibliographic universe—the one that we all visit from time to time—there are a number of different kinds of entities that are capable of entering into relations with one another. We might find it convenient to distinguish in some way between worlds, works, words, persons, and so on. It doesn't really make much difference whether we decide to treat these entities as substances that somehow exist separate from their properties, or simply as bundles of properties. Whatever system of fundamental categories of entities we settle on, we can also use it as the basis of a taxonomy of relations between entities. Depending on our purposes, we might want to distinguish (a) relations between works and people from (b) relations between works and other works, for instance.

Alernatively, we could look at properties of the relations themselves, rather than properties of the entities being related. For instance, we could distinguish

between kind/instance relations and whole/part relations, and so on. Lots of people have done this: Rebecca Green and Elaine Svenonius, for example (see, e.g., Green 2002; Svenonius 2000).

Let us take a look at the relation of *aboutness*. Say we have two documents, Doc 1 and Doc 2. If I decide that Doc 1 is about Subject A, what that amounts to is a judgment—an entirely subjective judgment made by me on a particular occasion—that Doc 1 has the property of being about Subject A, that Doc 1 is a member of the class of documents that share the property of being about Subject A, that Doc 1 instantiates Subject A ... All these are different ways of saying exactly the same thing.

There are many different ways of modeling that relation, and many different ways of visually depicting it, and we would use different modeling techniques depending on whether we were working within a philosophical framework, or a computer science framework, or a library and information science framework. We would come up with different models depending on whether we wanted to emphasize aboutness as a property, with subjects treated as attributes of documents, or aboutness as a relation, with subjects treated as classes of documents. But this diagram (Figure 1) tells the basic story. We can also say that Doc 1 and Doc 2 are *similar* in the sense that they share the same property or that they are members of the same class. Doc 1 and Doc 2 are not the same document, but they are the same *kind* of document. Another thing we can say is that, if it turns out that Subject B has exactly the same extension as Subject A does—in other words, if it turns out that all and only the documents that are about Subject A are about Subject B—then we can say that Subject A is the same subject as Subject B. They are not merely similar; they are identical. In fact, there is only one subject, not two.

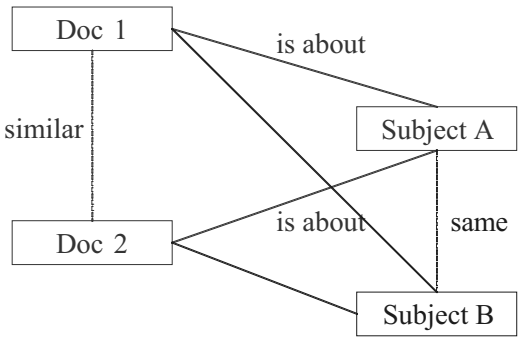

Figure 1. Aboutness.

Just as a quick digression, here is a representation of the relation of *work-instantiation* (Figure 2). The structure of the relation is exactly the same. If I decide that Doc 1 instantiates Work A, what that amounts to is a judgment—an entirely subjective judgment made by me on a particular occasion—that Doc 1 has the property of being an instance of Work A, that Doc 1 is a member of the class of documents that share the property of instantiating Work A. Again, these are just different ways of saying the same thing, and again we can also say that Doc 1 and Doc 2 are similar in the sense that they share the same property or that they are members of the same class. And again, if it turns out that Work B has exactly the same extension as Work A does—in other words, if it turns out that all and only the documents that instantiate Work A instantiate Work B—then we can say that Work A is the same work as Work B.

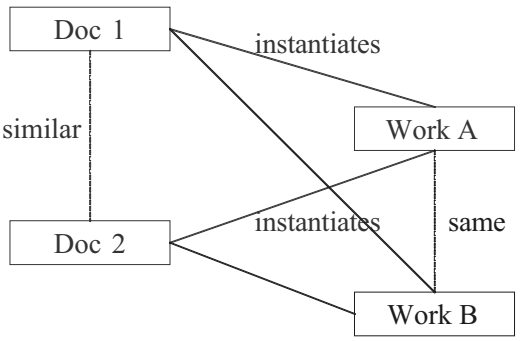

Figure 2. Work-instantiation.

Relevance works the same way (Figure 3). If I decide that Doc 1 is relevant to Subject A, what that amounts to is a judgment—an entirely subjective judgment made by me on a particular occasion—that Doc 1 has the property of being relevant to Subject A, that Doc 1 is a member of the class of documents that share the property of being relevant to Subject A. And if all and only the documents that are relevant to Subject A are relevant to Subject B, then we can say that Subject A is the same subject as Subject B.

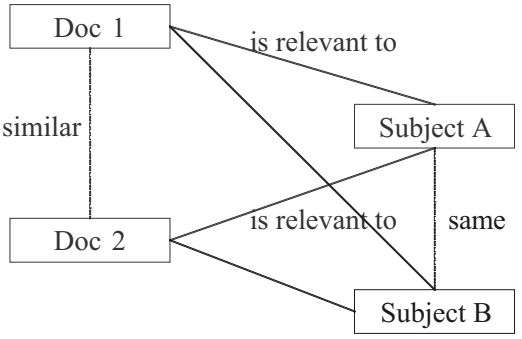

Figure 3. Relevance.

This is the sense in which aboutness and relevance—and, incidentally, work-instantiation—are equivalent in structure. An understanding of identity, and the difference between similarity and sameness, is obviously helpful in analyzing that structure.

Meanwhile, it is important to note the assumption here that subjects (and works) are not natural kinds. They are human artifacts—nominal kinds or social constructs. Subjects do not somehow inhere in documents. They are not properties of documents: they are properties of intentional acts. Judgments of aboutness, work-instantiation, and relevance are arbitrary and subjective. Whenever somebody—anybody—says that Document 1 is about Subject A, all that person means is "I currently think this document is a member of the class I currently call A" (cf. Wilson 1968).

5.0 Identity as the goal of KO

The main claim that I would like to make about the importance of identity for KO is not that an understanding of identity is helpful in analyzing the structure of aboutness and relevance. It is that there is a sense in which identity is actually the goal of KO.

There is a view—and it is a useful view, I think—that the aim of information retrieval system design is to *produce* identity, to produce matches or correspondences, between sets of aboutness judgments and sets of relevance judgments. If the system retrieves all and only those documents that are judged by the searcher to be relevant, then the system is successful. In this view, it is just the judgments made by the system, as to whether given documents are members of the classes named in the user's queries to the system, which are being characterized as aboutness judgments. In this account, the aim of indexing—i.e., the aim of making aboutness judgments—is to achieve consistency between the aboutness judgments made by indexer a with respect to subject x at time t_1, and the relevance judgments made by searcher b at time t_2.

And the aim of KO in general is to achieve consistency—to produce identity—between (i) the KO system designer's representation of reality, which basically amounts to the aggregate of the extensions of all subject classes and the relations between them, and (ii) the KO system user's model or image of the world. The challenge for KO is simply that there are many different views of the world. The challenge is not just that there is no single view, but that there at least six billion! So there can be no doubt: If KO systems are to work well, then they have to be dynamic and adaptive in whatever ways are productive of consistency between the system and each individual user.

I have talked about aboutness and relevance as different relation-types. These two relation-types are similar in the sense that they are themselves instances of the same broad class of relation-types: the class formed by relation-types whose own instances relate classes, kinds, sets, or types (on the one hand), and the instances of those classes or kinds, the elements of those sets, or the tokens of those types (on the other). There is another broad class of relation-types, and this one is made up of relation-types whose instances are class-to-class or set-to-set, such as the genus–species relation-type. This is the class of relation-types whose instances relate subjects in KO systems such as faceted, hierarchical classification schemes. In the example of such a scheme depicted in Figure 4, there is a broader–narrower relationship between Subject A1 and Subject A11, and another one between Subject A1 and Subject A12. Notice three things here: (a) Docs 2 and 3 are instances of subjects both in Facet A and in Facet B, but Doc 1 is an instance only of subjects in Facet A. That is okay, of course. (b) Docs 1 and 2 are instances of both Subject A1 and one of its two sub-classes, but Doc 3 is an instance only of Subject A1. That is okay, too. (c) Note that we are not saying anything about the semantic relations among the things or concepts that are represented by whatever terms are used as the labels for these Subjects. The relation between Subject A1 and A2 is simply a class-to-class relationship. The thing or concept represented by the term used as the label for the narrower class may be an instance of the kind that is represented by the term used as the label for the broader class, or it may be a subclass of the class that is represented by the broader term.

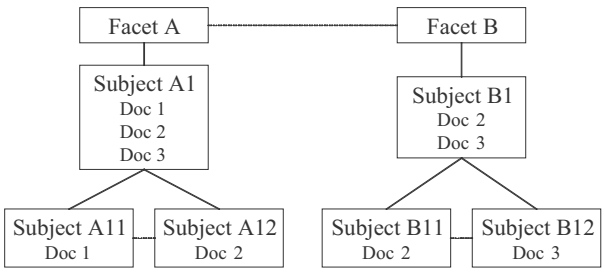

Figure 4. Inter-class relations.

These are the kinds of relations that have historically been put to use in faceted, hierarchical classification schemes. The question I want to ask at this point is this: Is this small set of relation-types—together with the rules of logic that constrain how particular

subject-classes may be instantiated in relation to other subject-classes—adequate for the representation of relationships between classes of documents about people? Obviously, I am not the first to ask this question. The work of people like Rebecca Green, Joan Mitchell, Hope Olson, Elaine Svenonius, and Barbara Tillett takes this question in all sorts of fascinating directions (see, e.g., Green 2008; Mitchell 2008; Olson 2002; Svenonius 2000; Tillett 2002). And I am not even going to attempt to answer this question here. But I do think that it is worth posing in the present context, because I think it is helpful to think about it as a special case of the general question of how well KO allows for the construction of system/user identity.

One thing to be made clear at this point is that some subjects of documents are the identities of the persons to whom the documents are relevant. As we saw earlier, one way in which different persons' views of the world vary is the way in which different persons have different views of their own and other people's personal and social identities. In fact, the phenomenon of personal identity is fairly complex, and the challenge for KO is the possibility that this complexity of identity relations cannot be represented by simple hierarchical class divisions.

Again, as we saw earlier, there are two central research questions in KO: the design question and the evaluation question. We can translate these general research questions into a couple of specific research questions that relate to KO's handling of identity. The first is the design question. How ought *identities-as-subjects*, and the relations between them, to be represented in a KO system? Again, the answer to this is that they should be represented in whatever way that evaluations tell us is best. But then there's the evaluation question: How do we evaluate? *How do we decide* how identities-as-subjects, and the relations between them, ought to be represented in a KO system?

Earlier, we also took a look at some of the criteria we can use to decide whether a particular representation of reality is any good or not. We said that it is possible to take a description-oriented view or a retrieval-oriented view of which are the most appropriate criteria to use. We identified justness among the extrinsic criteria on the description-oriented side, and talked about effectiveness as the most important of the retrieval-oriented criteria. We looked at some of the factors that affect levels of effectiveness. One of those factors was indexer–searcher consistency. Now, we can say that one component of any strategy that is intended to maximize the degree of consistency between categorizers—between indexers and searchers, for instance—is to ensure that categorizers' *self-identities* are *reproducible* in the KO system.

What does reproducible mean here? One way of clarifying this might be to draw up a kind of "bill of rights" for KO system users (see Table 4). Ask yourself, intuitively: Can our opportunities to do these things with KO systems be conceived as rights that should be protected like other basic human rights? Do I not have the right to find documents that are relevant to any one or any combination of my multiple personal identities—(i) as effectively, efficiently, and easily as I would find documents about any *other subject*, and (ii) as effectively, efficiently, and easily as *anyone* would find documents about any of *their* personal identities? Do I not have the right to use, and expect others to understand, my own vocabulary in communicating about identities-as-subjects—without hurting effectiveness, efficiency, or ease of retrieval in any way? And do I not have the right to describe identities-as-subjects, including my own, differently in different situations and at different times—again, without hurting effectiveness, efficiency, or ease of retrieval in any way?

I have the right ...
... to find resources that are relevant to any one or any combination of my **multiple** personal identities—as effectively, efficiently, and easily (i) as I would find resources about any other subject, and (ii) as anyone would find resources about any of their personal identities
... to use, and expect others to understand, **my own vocabulary** in communicating about identities-as-subjects—without hurting effectiveness, efficiency, or ease of retrieval
... to describe identities-as-subjects, including my own, **differently** in different situations and at different times—without hurting effectiveness, efficiency, or ease of retrieval

Table 4. A Bill of Rights for Autonomous KO Users.

6.0 Some challenging complexities

There are a number of complexities in this account of what it means to have an identity, and each of the complexities poses a challenge of a different kind for the designers of KO systems used to support the provision of access to documents about identities. I am going to take a brief look at five of these in turn. I label them as follows: multifacetedness of personal identities; individual differences in prioritization of

facets; intra-facet mixedness; intra-facet multidimensionality; and vagueness. Each of these ideas on its own is quite straightforward, but put them all together and you have quite a complex situation.

1. *Multifacetedness* of personal identities. Multifacetedness is quite simple. The idea is that every person may have multiple identities, in the sense that each person may simultaneously affiliate with multiple classes, each of which is defined by a property instantiating a different facet. So, for example, I might simultaneously identify with forty-somethings, with males, with Brits, and so on.
2. Individual *differences in prioritization* of facets. Different persons prioritize their affiliations in different ways, at different times. Different persons who have the same set of multiple affiliations may have different "defining characteristics," or identities that they consider to be the most important, or the most strongly influential on their decisions and actions. For example, a person who self-identifies as a middle-class, racially-mixed lesbian may self-identify most strongly as a middle-class person, or as a woman, or as a racially-mixed woman, or ...
3. Intra-facet *mixedness*. There is another sense in which a person may simultaneously or diachronically have multiple identities. This is the sense in which a person may affiliate with multiple classes defined by different properties in the same facet, so that for example somebody who self-identifies as racially mixed may self-identify with one racially-defined population at one time and another racially-defined population at different times, or even with multiple racially-defined populations at the same time. Here it is instructive to refer to Maria Root's "Bill of Rights for Racially Mixed People," which includes the rights to self-identify with more than one group of people, to self-identify differently at different times, to self-identify in ways that may be contrary to other people's expectations or wishes, and to use one's own vocabulary to communicate about one's multiraciality or mixedness (Root 1996).
4. Intra-facet *multidimensionality*. The idea here is simply that some facets are not unidimensional. Take sexual orientation, for instance. Is it really the case that all sexual orientations can be located on a single dimension with homosexuality at one pole, heterosexuality at the other, and bisexuality in the middle? Or should be thinking about sexual orientation as a two-dimensional thing, where different orientations can be located on a graph with degree of attraction to members of the same sex on one axis, and degree of attraction to members of a different sex on the other?
5. *Vagueness*. Fifthly and finally, we have the observation that the boundaries of the classes with which persons affiliate are themselves vague. These classes are not natural kinds; they are not classes whose definitions are somehow discoverable through scientific enquiry. They are nominal kinds or artifactual kinds whose memberships are determined over time by convention. Not only that, but these are classes whose members are identified as being members of those classes not because they share certain specified properties that are somehow known to all agents doing the identifying, but because they resemble, more or less closely, certain prototypes known from experience to have been identified by others as core exemplars of those classes.

Of course, a lot of work, theoretical and empirical, has been done in the last hundred years or so that has had the general effect of changing our understanding of how people categorize in general, and I would just like to throw in the name of Timothy Williamson here as an example of the work being done in mainstream philosophy on the implications of vagueness for traditional Aristotelian theories of logic that tend to rely on assumptions about the reality of binary dichotomies (Williamson 1994). Many voices in KO have been arguing for a long time that, if the design of KO systems is to improve, then we somehow have to start taking seriously these ideas about how categorization actually proceeds. Of course, Clare Beghtol, Hope Olson and others are the authorities here (see, e.g., Beghtol 2005; Olson 2002).

7.0 Conclusion: The "Third Way"

This is all a statement of what should happen—of what we might wish for. Of course, we all want to respect the rights of users. I should think it unlikely that anyone would want to start arguing that in fact we have got it all wrong when it comes to protecting basic rights and promoting the just KO system, and that KO systems would be better if they were unjust, and so on. I know I am preaching to the choir here.

But the central issue is this: We *want* our KO systems to represent all identities-as-subjects in a just manner that respects everyone's rights. But how *can* they—given the multifacetedness, and the multiprioritizability, and the mixedness, and the multidimensionality, and the vagueness of identities-as-subjects?

I will give you a clue about how I think we *cannot* do it. And that is through traditional library classification—that is, through hierarchies of classes that are divided up one dimension at a time. Or even through facet analysis, and this is where I very much agree with Birger Hjørland's comments in *his* paper in the latest issue of KO (Hjørland 2008).

I think we need to look at the very methods we use in our current systems to represent the relations between classes, and take a critical, questioning perspective on those methods, which asks: Is that all there is? I think there has got to be a third way, and it is probably a way that takes what we have learnt from IR approaches and from folksonomic approaches, and applies that—with the intention not of adding to the fuel of those age-old arguments about who should do the indexing (people or machines? experts or hoi polloi?), but with the intention of discovering new ways of adequately allowing for the representation of the multiprioritizability, the mixedness, the multidimensionality, the vagueness, as well as the multifacetedess of identities as subjects.

To emphasize: This is not about removing people from the process and replacing them with machines. It is not about rejecting vocabulary control. It is about advocating for a new kind of structure for representing the relations among documents. It is about encouraging people to look at different kinds of relationships from the ones that perhaps they are used to looking at, looking carefully at the similarities between those relationships, similarities in structure and similarities in role, and determining the implications of those discoveries of similarity for the design of representational structures. And it is about engaging seriously with the challenges for KO that are presented by analyses of identity and identity-forming processes.

8.0 Postscript

Finally, I just have a few remarks about the prospects for a fully-fledged philosophy of documentation or philosophy of knowledge organization. Several authors have recently picked up on Margaret Egan's social epistemology (Egan & Shera 1952), or at least the version of it promoted by Jesse Shera, as a good source of ideas about what the philosophical foundations of the information sciences are, or what they could be. The general idea seems to be that it is possible to move from an understanding of the social processes by which we most reliably acquire true or useful or relevant beliefs, to recommendation of the ways in which our information services can best support those processes. Don Fallis is one of the people who is doing a lot to show how this move works (see, e.g., Fallis 2006).

But Luciano Floridi and others have suggested that this kind of social epistemology cannot on its own provide a complete philosophy of information, since it focuses, obviously on knowledge acquisition processes and does not deal directly with metaphysical questions about the nature of information, the nature of documents, the nature with our interactions with those kinds of things and so on (Floridi 2002). Not only that but it does not deal directly with ethical questions about the kinds of justification that might be required of calls for the just information service or the diverse information service, and so on. In this context, maybe a metaphysics and an ethics of identity is just what is needed to augment the social epistemological approach to a philosophy of KO.

References

Abrams, Dominic. 2001. Social identity, psychology of. In Smelser, Neil J., and Baltes, Paul B., eds., *International encyclopedia of the social & behavioral sciences*. Amsterdam: Elsevier, pp. 14306-09.

Beghtol, Clare. 2005. Ethical decision-making for knowledge representation and organization systems for global use. *Journal of the American Society for Information Science and Technology* 56: 903-12.

Bilgrami, Akeel. 2001. Identity and identification: Philosophical aspects. In Smelser, Neil J., and Baltes, Paul B., eds., *International encyclopedia of the social & behavioral sciences*. Amsterdam: Elsevier Science, pp. 7148-54.

Cerulo, Karen A. 1997. Identity construction: New issues, new directions. *Annual review of sociology* 23: 385-409.

Egan, Margaret E., and Shera, Jesse H. 1952. Foundations of a theory of bibliography. *Library quarterly* 22: 125-37.

Fallis, Don. 2006. Social epistemology and information science. *Annual review of information science and technology* 40: 475-519.

Fidel, Raya. 1994. User-centered indexing. *Journal of the American Society for Information Science* 45: 572-6.

Floridi, Luciano. 2002. On defining library and information science as applied philosophy of information. *Social epistemology* 16: 37-49.

Forrest, Peter. 2006. The identity of indiscernibles. In Zalta, Edward N., ed., *Stanford encyclopedia of phi-*

losophy. Stanford, CA: Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University. Available http://plato.stanford.edu/entries/identity-indiscernible/.

Furner, Jonathan. 2007. Dewey deracialized: A critical race-theoretic perspective. *Knowledge organization* 34: 144-68.

Gallois, Andre. 2005. Identity over time. In Zalta, Edward N., ed., *Stanford encyclopedia of philosophy*. Stanford, CA: Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University. Available http://plato.stanford.edu/entries/identity-time/.

Green, Rebecca. 2008. Relationships in knowledge organization. *Knowledge organization* 35: 150-9.

Hardin, Russell. 2001. Identity: Social. In Smelser, Neil J., and Baltes, Paul B., eds., *International encyclopedia of the social & behavioral sciences*. Amsterdam: Elsevier, pp. 7166-70.

Heyes, Cressida. 2007. Identity politics. In Zalta, Edward N., ed., *Stanford encyclopedia of philosophy*. Stanford, CA: Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University. Available http://plato.stanford.edu/entries/identity-politics/.

Hjørland, Birger. 2008. What is knowledge organization (KO)? *Knowledge organization* 35: 86-101.

López-Huertas, María J. 2008. Some current research questions in the field of knowledge organization. *Knowledge organization* 35: 113-36.

Marcia, James E. 2001. Identity in childhood and adolescence. In Smelser, Neil J., and Baltes, Paul B., eds., *International encyclopedia of the social & behavioral sciences*. Amsterdam: Elsevier, pp. 7159-63.

Mitchell, Joan S. 2008. Groups of people in Table 1 and 305-306. *025.431: The Dewey blog*, August 1. Available http://ddc.typepad.com/025431/2008/08/since-late-2005.html.

Noonan, Harold. 2006. Identity. In Zalta, Edward N., ed., *Stanford encyclopedia of philosophy*. Stanford, CA: Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University. Available http://plato.stanford.edu/entries/identity/.

Olson, Hope A. 2002. *The power to name: Locating the limits of subject representation in libraries*. Dordrecht: Kluwer Academic.

Renear, Allen H., and Dubin, David. 2007. Three of the four FRBR Group 1 entity types are roles, not types. In Grove, Andrew, and Rorissa, Abebe, eds., *Proceedings of the 70th ASIS&T Annual Meeting* (Milwaukee, WI, October 19-24, 2007). Medford, NJ: Information Today.

Root, Maria P. P. 1996. A bill of rights for racially mixed people. In Root, Maria P. P., ed., *The multiracial experience: Racial borders as the new frontier*. Thousand Oaks, CA: Sage, pp. 3-14.

Smiraglia, Richard P. 2001. *The nature of "a work": Implications for the organization of knowledge*. Lanham, MD: Scarecrow.

Svenonius, Elaine. 2000. *The intellectual foundation of information organization*. Cambridge, MA: MIT Press.

Tennis, Joseph T. 2008. Epistemology, theory, and methodology in knowledge organization: Toward a classification, metatheory, and research framework. *Knowledge organization* 35: 102-12.

Tillett, Barbara B. 2002. Bibliographic relationships. In Bean, Carol A., and Green, Rebecca, eds., *Relationships in the organization of knowledge*. Dordrecht: Kluwer Academic, pp. 19-35.

Williamson, Timothy. 1994. *Vagueness*. London: Routledge.

Wilson, Patrick. 1968. *Two kinds of power: An essay on bibliographic control*. Berkeley, CA: University of California Press.