Proposal for a Classification of Classifications built on Beghtol's Distinction between "Naïve Classification" and "Professional Classification"

Elin K. Jacob

School of Library and Information Science, Indiana University Bloomington,
011 Wells Library, 1320 E. 10th St., Bloomington, IN 47405 USA <ejacob@indiana.edu>

Elin K. Jacob is an Associate Professor, Director of the Doctoral Program, and Director of the Semantic Web Lab at the School of Library and Information Science, Indiana University Bloomington. She received her Ph.D. in Library Science from the University of North Carolina-Chapel Hill and teaches courses in representation and organization, metadata and ontologies. Her current research interests include classification theory, alignment of ontologies, interoperability of metadata schemas and the analysis of social bookmarking vocabularies

Jacob, Elin K. **Proposal for a Classification of Classifications built on Beghtol's Distinction between "Naïve Classification" and "Professional Classification".** *Knowledge Organization, 37(2)*, 111-120. 27 references.

ABSTRACT: Argues that Beghtol's (2003) use of the terms "naïve classification" and "professional classification" is valid because they are nominal definitions and that the distinction between these two types of classification points up the need for researchers in knowledge organization to broaden their scope beyond traditional classification systems intended for information retrieval. Argues that work by Beghtol (2003), Kwaśnik (1999) and Bailey (1994) offer direction for the development of a classification of classifications based on the pragmatic dimensions of extant classification systems.

1.0 Introduction

Over her distinguished career, Claire Beghtol has made significant contributions to the field of knowledge organization. While her work has ranged broadly, it has been constrained by her focus on classification, be it from an historical, theoretical, linguistic, or ethical perspective. One of Beghtol's more recent contributions to knowledge organization in general and classification research in particular is the distinction she has made between "professional classifications" that support information retrieval and "naïve classifications" that facilitate knowledge creation (Beghtol 2003). Although her paper outlining this dichotomy was generally well received, it engendered a heated response from Hjørland and Nicolaisen (2004), who raised particularly strenuous objections to Beghtol's use of the term "naïve" in association with classification.

In this paper, I argue that the objections raised by Hjørland and Nicolaisen were not simply unfounded but were, in fact, a disservice to knowledge organization in that they drew attention away from Beghtol's (2003, 65) primary argument: "that it is important to study the relationships among classification systems developed for different purposes." Classification research in knowledge organization has been unnecessarily limited (and limiting) by a single-minded focus on classification schemes used for the retrieval of resources. To establish the relevance of knowledge organization for the broader scholarly community—for both the sciences and the humanities—research in

knowledge organization must extend its sights beyond the purview of information retrieval systems and develop a systematic body of knowledge that addresses all aspects and applications of classifications. In short, research in knowledge organization must work toward the development of a classification of classifications and the establishment of a science of classification.

2.0 The case for "naïve" and "professional" classifications

The Eighth International ISKO Conference was held in London in the summer of 2004. I was in the audience when Beghtol gave her presentation on naïve classification systems, and I distinctly remember a member of the audience congratulating her for yet another perceptive and illuminating contribution to classification research. Unfortunately, the full significance of Beghtol's contribution, both in her London presentation (2004a) and in her earlier paper (2003), was overshadowed by Hjørland and Nicolaisen's (2004) criticism of the distinction Beghtol posed between "professional classifications" and "naïve classifications."

In her 2003 paper, Beghtol (65) had observed that, within knowledge organization, "the conventionally accepted purpose of classification systems for … information retrieval has been to reveal knowledge … that has already been stored in documents" and that "the use of classification for knowledge discovery has been less studied [in knowledge organization] than classification for knowledge discovery." Beghtol's argument here was not that researchers should more actively pursue the creation of new knowledge through "mining" existing classification systems to identify hitherto unrecognized connections. Rather, as she clarifies in her response to Hjørland and Nicolaisen, Begthol (2004b, 62) was suggesting an important distinction between "professional classifications"—including the major disciplinary classifications as well as traditional bibliographic classifications—and "scholarly classifications for small areas of new knowledge within [existing] major disciplines." This distinction, she contended, was one of purpose: The purpose of scholarly classifications (i.e., naïve knowledge discovery classifications) "is significantly different from the purpose of the major disciplinary classifications" and "strongly influence[s] the content and extent of the two kinds of classifications" (Beghtol 2004b, 62).

The purpose of professional information retrieval classifications, and especially of bibliographic classifications, is, as the name implies, the provision of access to and the retrieval of information resources. With respect to disciplinary classifications, Mayr (1982, 148) observes that the practical purpose of classifications is to "serve as an index to an information storage and retrieval system. … Such a classification is automatically the key to the information stored in it." But the purpose of naïve knowledge discovery classifications is, as Beghtol (2003, 65) argues, "to enhance domain knowledge for the pursuit of scholarly activity and research" in a given discipline and to "support a scholarly environment in which new questions are expected to be asked of primary research materials."

Hjørland and Nicolaisen (2004, 55-56) reacted to Beghtol's original argument based on a very personal interpretation of the juxtaposition between "naïve classifications" and "professional classifications." Their failure to look beyond an everyday understanding (and the associated pejorative connotations) of the terms "naïve " and "professional" led them to raise the objection that, in her use of the phrase "naïve classifications:"

> Beghtol describes the classifications developed by scholars as 'naïve' while she describes the classifications developed by librarians and information scientists as 'professional.' We fear that this unfortunate terminology is rooted in deeply anchored misjudgments about the relationships between scientific and scholarly classification on the one side and LIS classifications on the other .... Classifications are produced in both [sic] the sciences, the social sciences and the humanities and they are important for how 'information retrieval classifications' should be designed. They are not 'naïve' compared to library classification, [sic] if anything it is the other way round.

I must admit that I, like Hjørland and Nicolaisen, had interpreted Beghtol's labels in light of my own colloquial understanding of the terms "naïve" and "professional." By focusing on the labels themselves, I had equated "naïve classifications" with "folk classifications;" and, in so doing, I had wrongly imbued these phrases with a pejorative import that was not only inappropriate but, more importantly, irrelevant to Beghtol's argument.

Cohen and Nagel (1934, 225) have observed that "serious blunders in reflective thinking occur because the meaning that a word has in some context is replaced, without the fact being noticed, by an allied but different meaning." Thus, according to *WordNet*

(Miller 2009) synonyms for the adjective "naïve" include, among others, "primitive," "inexperienced," and "uninstructed." These very familiar and commonplace terms share the connotation of "simplicity" or "lack of experience;" more importantly, as Cohen and Nagel (1934, 225) suggest, they are infused with "strong emotional associations and carry penumbras of suggested meanings which obstruct the process of rigorous deduction." By applying this everyday understanding of "naïve" to the phrase "naïve classification," Hjørland and Nicolaisen have effectively stripped it of the ability to communicate Beghtol's original intent. More importantly, they have imbued "naïve classification" with a meaning that obfuscated the core of her argument.

A close reading of Beghtol's (2003) paper reveals that she neither claims that classifications produced in the sciences and the humanities are "naïve" nor insinuates that scholarly classifications are somehow less "professional" than classification schemes used for information retrieval. She is very careful to explicitly define the referents of "naïve classification" and "professional classification," both in the body of her paper and in the abstract (Beghtol 2003, 64):

> In this paper, classifications for information retrieval are called "professional" classifications because they are devised by people who have a professional interest in classification, and classifications for knowledge discovery are called "naïve" classifications because they are devised by people who have no particular interest in studying classification as an end in itself.

To eliminate any ambiguity or popular nuances that might be associated with the terms "naïve" and "professional," Begthol has provided her readers with explicit definitions that establish, beyond question, the meaning of the labels "naïve classifications" and "professional classifications." These are nominal definitions and, as such, they cannot be challenged because they are neither true nor false (Cohen and Nagel 1934, 229). Whether or not the individual reader agrees with Beghtol's choice of labels, these definitions are not open to criticism based on terminology; as nominal definitions, they serve simply as a convenient shorthand that assures agreement between Beghtol and her audience as to the specific referents of the terms "naïve classification" and "professional classification." Beghtol has left no room for the "emotional associations" (Cohen and Nagel 1934, 229) that lead to "serious blunders in reflective thinking" (Cohen and Nagel 1934, 225). Obviously, the criticisms of Hjørland and Nicolaisen are unfounded.

3.0 Categorization, classification, and theory

Bailey (1994, 1) points out that classification is a central feature of everyday life: "It is so ubiquitous that not only do we generally fail to analyze it, we often even fail to recognize its very existence." Given widespread agreement that the primary characteristics of classification are exhaustivity and mutual exclusivity, Jacob (1992, 2004) has argued that it is not classification but the cognitive process of categorization that is, in fact, ubiquitous.

Categorization involves the apprehension of similarity across entities and the subsequent division of the experiential world into groups (i.e., categories) whose members share some perceptible similarity within a given context. Zerubavel (1991, 5) observes that the individual finds order and meaning in the environment by imposing boundaries—by splitting and lumping objects of experience into groupings that represent distinct "islands of meaning." It is just this ability to manipulate the environment through the creation of categories that allows the individual to forge new relationships and develop new knowledge whose value exceeds the simple grouping of objects: The flexibility and plasticity of the categorization process support the individual's ability to create and modify the informational content of a category as a function of context, purpose or experience.

The acquisition of new knowledge is dependent on the ability to create new categories through the discovery of new patterns and new relationships. With the accumulation of ever more specialized knowledge and the appearance of ever more specialized disciplinary domains, these categories become formalized as well-defined classes (Jacob, 1994) As these experientially-derived categories evolve into well-defined, domain-specific classes, they are frequently organized within formal classificatory structures that establish relationships between classes and facilitate the sharing of extant knowledge, thereby assuring that disciplinary knowledge is consistent not only across individuals in the discipline, but also across time. In this process, however, classes may lose their ability to respond to new patterns of similarity; and, as knowledge in the discipline advances, scholars are forced to review, revise and ultimately reconstruct both the classes themselves and their classificatory relationships.

Although there are obvious similarities between the cognitive process of categorization and the formal

process of classification, there is an important distinction: while the process of categorization is flexible and draws non-binding associations between entities based on the recognition of immediate similarity, the process of classification is rigorous in that an entity either is, or is not, a member of one particular class. The process of classification entails systematic assignment of all members of a collection to a set of mutually exclusive and non-overlapping classes, generally arranged within a formal structure that represents relationships between these classes. While a classification may assume one of several different relational forms (e.g., hierarchy, tree, paradigm (Kwaśnik 1999)), the class structure must demonstrate two fundamental characteristics: It must be both jointly exhaustive and mutually exclusive. As Bailey (1994, 3) explains, "if *N* persons are to be classified, there must be an appropriate class for each (exhaustivity), but only one correct class for each, with no case being a member of two classes (mutual exclusivity). Thus, there must be one class (but only one) for each of the *N* persons."

Classification creates order through the assignment of entities to a set of mutually exclusive and well-defined classes according to a specific theory or established set of principles. This order may be both artificial and arbitrary: artificial in that it is an artifact for establishing order within a domain and arbitrary in that it is governed by one perspective—one worldview or one theory—to the exclusion of all others. More importantly, the possibility of introducing order within an empirical domain assumes that knowledge can be both defined and arranged—that knowledge is "at the same time describable and orderable" (Foucault 1970, 158). In this way, classification not only creates order but is simultaneously shielded from the vagaries of meaning inherent in everyday language. Classification as a well-defined language ensures the stability of disciplinary nomenclature and facilitates the transmission of knowledge precisely because it is independent of context and thus conveys meaning directly and immediately.

The essential observation, however, must be that the process of classificatory organization and identification is necessarily carried out within a theoretical framework. Most people would agree that the purpose of science is to provide an understanding of the environment by describing things and events, explaining why and how those things and events happen, and offering a sense of control over the future by facilitating the prediction of events and outcomes. The first of these purposes—that of description—is perhaps the simplest to satisfy, because, as Reynolds (1971, 4) argues, "any set of concepts can be used to organize and classify."

Reynolds (1971, 10-11) has identified three predominant forms of scientific theory and evaluated each of these forms in light of the three purposes of science. The *set-of laws* approach conceives of scientific theory "as a set of well-supported empirical generalizations or 'laws.'" The *axiomatic* approach draws upon the theory of mathematics, conceiving of scientific theory "as an interrelated set of definitions, axioms, and propositions … derived from the axioms." The *causal process* approach conceives of theory "as a set of descriptions of causal process." Reynolds (1971, 83-114) has demonstrated how each of these theoretical forms clearly achieves the first purpose of science—that of organizing and classifying "things."

The primary problem generally associated with organizing and classifying phenomena is determining which of a plethora of dimensions would be most useful: one could organize and classify a group of individuals by height, weight, age, gender, educational level, marital status, etc. In practice, however, this problem is minimized by theory: If the theory postulates that income is correlated with educational level, it would be a waste of time and intellectual effort to record the height of each individual in a sample population. As Cohen and Nagel (1934, 249) have observed, "the kind of order, as well as the kinds of elements we look for, is determined by the nature of the problem [or theory] which generates the inquiry. What is an adequate answer to one question will not, in general, be adequate to another."

The centrality of classification and organization in the scientific process is made abundantly clear in the general suggestions Reynolds has offered to social scientists: "Any ideas proposed as potential scientific knowledge … should be as clear and as explicit as possible.… All concepts should be clearly defined and … relationships between concepts in statements should be made *explicit*" (Reynolds, 1971, 160; emphasis in original). In other words, theories should be constructed in such a way that they identify and define the concepts and relationships that will facilitate creation and organization of a system of classification.

Reynolds (1971, 160) has pointed out that the advance of scientific knowledge is, of necessity, an iterative process: theories are developed and tested, in part by classifying and organizing their component concepts, and are subsequently revised or even rejected, with the theories ultimately emerging via this process becoming "progressively more useful for the

purposes of science." Thus, the role of classification in the scientific process must be understood as "more than a catalog of neatly arranged and accessible facts" (Altman 1967, 64). Classification is necessarily a central feature of theory creation and evaluation: For example, a classification must be able to highlight deficiencies in its own organization or to identify areas where its internal distinctions are either too general or too specific.

Classification plays an important role not only in the creation of new theories, but also in the identification and development of new areas of scholarly endeavor. As Altman (1967, 64) has observed, classification can "lay the groundwork for new theory and point to new areas of study." And Cohen and Nagel have called attention to "the fact that the process of classifying things really involves, or is a part of, the formation of hypotheses as to the nature of things." But they have also cautioned that "the process of scientific classification is much more groping and less formal than [its] rules would suggest" (1934, 242); and they have underscored the "groping" nature of scientific endeavor, especially in the early stages of disciplinary development (1934, 243):

> The basis of classification depends on the discovery of some significant traits, significant in the sense that on the basis of the traits the subject matter can be organized into a system. Such traits, however, are only slowly discovered, and cannot be determined on formal grounds alone. All sciences in their early days are classificatory, and almost any arbitrary scheme of grouping objects may be tentatively adopted in the interest of mastery of the subject matter.

While classifying and organizing a set of phenomena is obviously a core purpose of scientific knowledge, it is a buttress for and not a stumbling block to the creation of theory, the formulation of new scholarly ventures, and the advance of science. As Reynolds (1971, 83-114) has established, each of the three approaches to theory in the social sciences—the set-of laws, axiomatic, and causal process approaches—facilitates the purpose of organizing and classifying phenomena that is the basis both for theory creation and development and for the identification of new research areas.

It is quite possible that research leading to a systematic classification of classifications (or classifications of classification) could contribute to the development of scientific knowledge in the social sciences and the humanities by identifying the appropriate structures, significant advantages, and noteworthy disadvantages of the various approaches to classification. Kwaśnik (1999) has provided a starting point for this research agenda; but any such venture will not be successful until scholars in knowledge organization are willing to acknowledge that information retrieval is not the singular function of classification—that there are, in fact, multiple functions served by classification systems. We must be willing to move outside the comfort zone of bibliographic classification and begin to study classification systems and classification practices "in the wild" (Beghtol 2003, 65; Jacob 2001, 78). If we are to contribute to the advance of scientific knowledge, we cannot bury our theoretical heads in the sand of bibliographic classification. We must revisit, revise, even reconstruct classification theory in light of recent advances in science and the unprecedented changes in the knowledge environment.

4.0 Classification, taxonomy and typology

Before addressing the question of possible dimensions that could be applied in the development of a classification of classifications, the question of terminology must be addressed.

Broughton (2004, 96) has defined a classification scheme "as a set of classes organized in a systematic fashion to show the relationships between them" and a class as "a set whose members share some common feature." Although Broughton's definition is ostensibly situated within the traditions of bibliographic classification, the lack of specificity and the very open-ended nature of this definition does not contribute to the agenda of creating a classification of classifications. The definition offered by Shera (1965, 99) is more focused. He has defined classification—bibliothetical classification—as:

> A list of terms which are specifically and significantly different each from the other, capable of describing the subject content of books, inclusive of all knowledge, infinitely hospitable, in an arrangement that is linear, unique, and meaningful, and which when applied to books, usually, though not necessarily, through the medium of a notation, results in their arrangement on the shelves according to the logical principle that inhere in the schematism.

Obviously, Shera's definition of classification is an ideal—even a utopian ideal—and, as such, it aptly il-

lustrates the criticism of W. Stanley Jevons (1887, 715; quoted in Shera 1965, 99) that classification in libraries is "a logical impossibility." More importantly, its list of criteria—and especially the requirements that a classification must be "inclusive of all knowledge" and generate a linear ordering "on the shelf"—limit classification, as Shera has defined it, to physical resources.

The definitions of Broughton and Shera are firmly situated within library science and the retrieval of information resources. Eschewing both the confines of traditional bibliographic classification and the ambiguity frequently found in definitions of classification (e.g., Broughton (2004, 295) has offered three distinct definitions of classification, all in terms of process), Bliss (1929, 143) has limited his definition to the arrangement of classes "in some order according to some principle or conception, purpose or interest, or some combination of such;" and he has defined the verb to classify as meaning "primarily to make, or conceive, a class, or classes, … [and] to arrange classes in some order or to relate them in some system according to some principle or conception, purpose or interest." This is, of course, a radical simplification of Bliss's description of classification and does not take into account his extended discussion of class or class-concepts (1929, 118-133). While these discussions would be of interest in the formulation of a classification of classifications, what is of value here is Bliss's emphasis on the essential role of "some principle or conception, purpose or interest" in devising an arrangement of classes. By moving beyond the confines of universal bibliographic classification, Bliss has been able to address classification as a general (and potentially universal) process of creating a set of related classes. And his focus on the role of "principle or conception, purpose or interest" (Bliss 1929, 143) underscores Beghtol's (2003) emphasis on the purposes for which naïve classifications are devised.

In *Sorting Things Out*, Bowker and Star move even further from the consideration of information retrieval, explicitly situating their definition of classification within the material world of experience. They define classification as "a spatial, temporal, or spatiotemporal segmentation of the world … a set of boxes (metaphorical or literal) into which things can be put to then do some kind of work" (Bowker and Star 1999, 10). They do concede that an ideal classification system would demonstrate certain formal properties—the operation of classificatory principles, mutual exclusivity, and exhaustivity or completeness. However, in their investigation of classification in the lived world, Bowker and Star adopt a very open-ended definition of classification: "anything consistently called a classification *and treated as such* can be included in the term" (Bowker and Star 1999, 13; emphasis in original).

Obviously, a research agenda seeking to develop a classification of classifications must build on a working definition of classification that is broader and more inclusive than Shera's definition of universal bibliographic classifications; but Bowker and Star's definition of classification as a "a set of boxes" is too broad and does not provide the definitional constraints that would distinguish classification from categorization. And the definitions offered by Broughton, like the willingness of Bowker and Star (1999, 13) to accept as an instance of classification "anything consistently called a classification *and treated as such*", are simply too ambiguous and provide no conceptual basis for determining what is and what is not a classification.

Given the inherent bias of definitions of classification developed within knowledge organization—definitions that naturally lean toward consideration of classification as a tool for information retrieval—it is appropriate to consider the definitions of those who are involved, either theoretically or practically, with the use of classification in data analysis, theory development and theory refinement.

Cohen and Nagel (1934, 223) have distinguished between division as the "breaking up" of a genus into its various species and classification as "the grouping of individuals into classes, and these classes into wider ones," where a class is defined as "a group of individuals each having certain properties in virtue of which they are said to be members of the class" (Cohen and Nagel 1934, 122). They have offered three "rules" that apply to both division and classification (i.e., exhaustivity, mutual exclusivity, and systematic grouping/division based on a single principle) with the added caution that these rules "are of little help in practice. They express an ideal rather than state a method. Moreover, the ideal is inadequate for a well-developed science: it is more suitable to sciences in their infancy" (Cohen and Nagel 1934, 242).

Cohen and Nagel approached classification from the perspective of philosophy. In contrast, Bailey (1994) has introduced the perspective of a social scientist, viewing classification as a research tool. He has defined classification "as the ordering of entities into groups or classes on the basis of their similarity … [seeking] to minimize within-group variance, while maximizing between-group variance. … By maximiz-

ing both within-group homogeneity and between-group heterogeneity, we make groups that are as distinct (nonoverlapping) as possible, with all members of a group being as alike as possible" (Bailey 1994, 1). In light of this definition, Bailey (1994, 3) has stipulated only a single rule or criterion that would constrain the process of classification: "the classes formed must be both exhaustive and mutually exclusive."

By combining the definitions provided by Cohen and Nagel, Bailey, and Bliss, we can forge a tentative definition of classification:

> Classification is the systematic process of grouping of a population of entities into a set of well-defined and mutually exclusive classes that minimize within-class variability and maximize between-class variability and of arranging the resulting classes in a formal structure that demonstrates relationships between classes according to some non-trivial principle or purpose.

This is, of course, a very preliminary definition of classification. It does not address Altman's (1967, 51) contention that classification is more appropriately conceived of "as 'selective study of research relationships' rather than systematic and comprehensive organizations." But it may very well offer a working foundation for a classification of classifications. For example, it provides for a definition of a classification as the product of the process of classificatory grouping. What this definition does not address, however, and what must be accounted for before it can be accepted as the most definitive definition of classification, is the relationship of classification—of a classification—to the related terms typology and taxonomy.

Bailey (1994, 4) has defined typology as "another term for classification." According to Bailey, however, a typology has two characteristics that distinguish it from a generic classification: "A typology is generally *multidimensional* and *conceptual*" (1994, 4; emphasis in original). In line with this emphasis on the multidimensional nature of typologies, Lewis-Beck (1994, v) has similarly observed that the "construction of a typology requires conceptualization along at least two dimensions." While a typology is inherently conceptual, a taxonomy is empirical. Bailey (1934, 6) has defined *a taxonomy* as "a classification of empirical entities … [that is] often (but not always) hierarchical (as in family, genus, species) and evolutionary." Again, Lewis-Beck (1994, v) has elaborated on Bailey's definition, pointing out that taxonomies are developed "empirically, rather than conceptually, with the goal of classifying cases according to their measured similarity on observed variables."

While these are very preliminary definitions and limited to a single source, it is possible that typology and taxonomy, like traditional bibliographic classification, are "types" of classification; and investigation of their use in the development of theory and the practice of research would contribute to the development of a classification of classifications.

5.0 Developing a classification of classifications

There are multiple avenues that could be explored in the development of a classification of classifications. As Cohen and Nagel (1934, 243) have observed, any such development must necessarily depend on the identification of significant, non-trivial dimensions that will facilitate construction of a meaningful relational structure; but these dimensions—the core traits of a broad conception of classification—"are only slowly discovered, and cannot be determined on formal grounds alone." Cohen and Nagel (1934, 243) have qualified the need for identification of "significant traits" when they allow that, for young sciences, which are necessarily classificatory, any arbitrary scheme of grouping objects may be tentatively adopted in the interest of mastery of the subject matter.

The implications of this for a proposed science of classification—a research agenda leading to the development of a classification of classifications—is the need for critical investigation of extant classifications, their development and their governing principles in light of their intended purpose or use. For example, Hjørland and Nicolaisen (2004, 55) have lauded the Mendeleev's periodic table of chemical elements: "The most respected and recognized of all scientific classifications is without much doubt the Periodic Systems of the Element in chemistry and physics." But Altman (1967, 56) has pointed to the "many irregularities and assymetries" in the periodic table. Thus, for example:

> the column arrangement according to metallic properties and organization of certain rare elements is not completely consistent or orderly. Thus, while useful and praised for nearly a century, the periodic table has not overcome certain inconsistencies and flaws. This, coupled with the extensive prior history of attempts to build a chemical classification system can be too easily

> ignored by those attempting to use Mendeleev as a golden standard.

The moral here must be that any exploration of extant classifications, their development and their principles cannot rely on tradition or reputation but must be truly unbiased and analytical in its investigation.

According to Reynolds (1971, 163), the ultimate problem facing the development of scientific knowledge in the social sciences has not been a lack of theory but "that much of what is written fails to meet the most fundamental criterion for a scientific body of knowledge:"

> the major factor that thwarts the development of a scientific body of knowledge of social and human phenomena is the character of the social scientists themselves—problems within the social scientists, not within the phenomena. Two major deficiencies are lack of clarity in theoretical writings and ignorance about what scientific knowledge should look like and how it is created.

Reynolds's observation underscores both the need for thorough assessment of the terminology of classification, relevant to classification broadly conceived, and for a methodical and critical assessment of the criteria governing the development and implementation of classifications.

Kwaśnik (1999) has discussed the characteristics, advantages and disadvantages of different classificatory forms and processes, providing potential dimensions that could be used in the development of a classification of classifications. There are, for example, the structural forms that she describes—hierarchies, paradigms and trees—and the various characteristics of these different forms; either of these could be used as a starting point for investigating the dimensions of classification and the distinctions (or the lack thereof) between different classificatory forms. Altman (1967, 59-66), too, has identified a variety of "choicepoints" as well as criteria for assessing the adequacy of classifications that could also provide dimensions for investigation of classifications.

One obvious starting point for this endeavor is Beghtol's identification of the purposes of naïve knowledge discovery classifications. Beghtol (2003, 66) has pointed out that naïve classifications are not created in a void, but place certain constraints on those who would develop them: "To support scholarly activities and research in some domain, naïve classifications require their creators to state a purpose for their classificatory endeavors and to encourage new ways of viewing the material(s) of the discipline(s) involved." To illustrate her argument, Beghtol (2003, 66) provides a list of "specific purposes" based on the five instances of naïve classification that she uses in the paper:

- Discover gaps in knowledge
- Fill gaps in knowledge
- Reconstruct historical situations and evidence
- Facilitate integration and communication of findings; and
- Suggest revisions or amplifications of accepted classifications.

An investigation of specific purposes would address Bliss's (1929, 143) emphasis on the "why" of classification: the "principle or conception, purpose or interest, or some combination of such" according to which classes are created and arranged.

Another avenue that might be productive in identifying dimensions appropriate for a classification of classifications has been provided by Bailey. Building on his three-level model of measurement (Bailey 1984; 1986), Bailey (1994, 3) has argued that the process of classification can incorporate three levels of analysis: "the conceptual (where only concepts are classified), the empirical (where only empirical entities are classified), and the … operational or indicator level … where [the conceptual and the empirical] are combined." According to Bailey (1994, 31), the conceptual level of classificatory analysis involves the deductive construction of a typology of conceptual types, real or imaginary, and does not include identification of empirical entities. Thus, a conceptual typology might have theoretical relevance, but there would be no practical implementation. At the opposite end of the conceptual-material continuum is the empirical level of classificatory analysis. The empirical level results in an inductive classification consisting of "a purely empirical taxonomy with no theoretical counterpart…. The basic strategy is simply to measure the empirical cases and group them by similarity … [into] empirically derived classes … that might have little or no real theoretical value or conceptual importance" (Bailey 1994, 31).

The combined or operational level generates the most common or "everyday" form of classification and is arrived at "by mapping both the conceptual and empirical levels into the third level … [and is] frequently confused with, or merged with, one of

these other levels" (Bailey 1994, 31). Bailey (1994, 32) described two approaches to arriving at the operational level of classification. The first, which he called "the classical strategy," is deductive; it is formed by first constructing a conceptual or theoretical model and then identifying empirical entities to populate the conceptual framework. The second approach, which Bailey does not name but which might be dubbed "the grounded theory strategy," is inductive. Because the process of constructing a classification according to the grounded theory strategy involves "clustering" a set of empirical entities and then devising labels for them, such a classification could feasibly be developed using machine clustering techniques.

Examples of the conceptual and operational levels of classificatory analysis—examples of conceptual and operational classifications—are relatively easy to identify. For example, the data-information-knowledge-wisdom pyramid would be an example of a purely conceptual typology, as also the "periodic table of expertises" constructed by Collins and Evans (2007, 14). Operational classifications are frequently exemplified in ethnographic research, which begins with the collection and grouping of empirical data, with descriptive labels subsequently assigned to the classes in the resulting taxonomy (Bailey 1994, 79). It seems obvious that, in most cases, enumerative bibliographic classifications are likely to be operational classifications constructed according to the classical strategy; in contrast, faceted classifications are likely are likely to be operational classifications constructed according to the grounded theory strategy. In any case, Bailey's notion of the three levels of analysis offers yet another possible set of dimensions that could be relevant in developing a classification of classifications.

6.0 Conclusion

By drawing a clear distinction between "professional classifications" and "naïve classifications," Beghtol's exposition of naïve knowledge discovery classifications constitutes a significant contribution to knowledge organization, not only for extending the understanding of classification but, more importantly, for bringing to the fore the need for knowledge organization to push beyond the limits of information retrieval. Beghtol (2003, 65) has argued "that it is important to study the relationships among classification systems developed for different purposes." In so doing, she has enlarged the scope of classification research and scholarly endeavors to include all classification; she has highlighted the need for knowledge organization to address the relevance of classification and classificatory structures for the conduct of research and the advance of scholarly knowledge; and she has dramatically expanded the potential theoretical and empirical influence of the field of knowledge organization.

If knowledge organization is to make a significant contribution to the advance of scientific knowledge, scholars and researchers in knowledge organization must work toward a more comprehensive understanding of classification through the development of a classification of classifications and, ultimately, realization of a science of classification.

References

Altman, Irwin. 1967. Choicepoints in the classification of scientific knowledge. In Indik, Bernard P. and Berrien, F. Kenneth, eds., *People, groups, and organizations*. New York: Teachers College Press, 47-69.

Bailey, Kenneth D. 1984. A three level measurement model. *Quality and quantity* 18: 225-45.

Bailey, Kenneth D. 1986. Philosophical foundations of sociological measurement: a note on the three level model. *Quality and quantity* 20: 327-37.

Bailey, Kenneth D. 1994. *Typologies and taxonomies: an introduction to classification techniques*. Thousand Oaks, Ca.: Sage.

Beghtol, Claire. 2003. Classification for information retrieval and classification for knowledge discovery: relationships between "professional" and "naïve" classifications. *Knowledge organization* 30: 64-73.

Beghtol, Claire. 2004a. *Naïve classification systems and the global information society*. Presented at the Eighth International ISKO Conference 13-16 July 2004 London, UK. Available: http://www.docstoc.com/docs/11322155/UNIVERSAL-CONCEPTS-CULTURAL-WARRANT-AND-CULTURAL-HOSPITALITY

Beghtol, Claire. 2004b. Response to Hjørland and Nicolaisen. *Knowledge organization* 31: 62-63.

Bliss, Henry Evelyn. 1929. *The organization of knowledge and the system of the sciences*. New York: Henry Holt.

Bowker, Geoffrey C. and Star, Susan Leigh. 1999. *Sorting things out: classification and its consequences*. Cambridge, MA: MIT Press.

Broughton, Vanda. 2004. *Essential classification.* London: Facet.

Cohen, Morris R. and Nagel, Ernest. 1934. *An introduction to logic and scientific method*. New York: Harcourt, Brace.

Collins, Harry and Evans, Robert. 2007. *Rethinking expertise*. Chicago: Chicago University Press.

Foucault, Michael. 1970. *The order of things: an archaeology of the human sciences*. New York: Vintage Books.

Hjørland, Birger and Nicolaisen, Jeppe. 2004. Scientific and scholarly classifications are not "naïve": a comment to Beghtol (2003). *Knowledge organization* 31: 55-61.

Jacob, Elin K. 1992. Classification and categorization: drawing the line. In Kwaśnik, Barbara H. and Fidel, Raya, eds., *Advances in classification research, 2*. Washington, DC: Learned Information, pp. 67-83.

Jacob, Elin K. 1994. Classification and crossdisciplinary communication: breaching the boundaries imposed by classificatory structure. In Albrechtsen, Hanne and Oernager, Susanne, eds., *Knowledge organization and quality management: advances in knowledge organization, 4*. Frankfurt/Main: Indeks Verlag, pp. 101-108.

Jacob, Elin K. 2001. The everyday world of work: two approaches to the investigation of classification in context. *Journal of documentation* 57: 76-99.

Jacob, Elin K. 2004. Classification and categorization: a difference that makes a difference. *Library trends* 52(3): 515-40.

Jevons, W. Stanley. 1887. *The principles of science*. London: Macmillan.

Kwaśnik, Barbara H. 1999. The role of classification in knowledge representation and discovery. *Library trends* 48(1): 22-47.

Lewis-Beck, Michael S. 1994. Series editor's introduction. In Bailey, Kenneth D., *Typologies and taxonomies: an introduction to classification techniques*. Thousand Oaks, Ca.: Sage, pp. v-vi.

Mayr, Ernst. 1982. *The growth of biological thought: diversity, evolution, and inheritance*. Cambridge, Ma.: Harvard University Press.

Miller, George A. 2009. *WordNet*. Available: http://wordnet.princeton.edu

Nicolaisen, Jeppe and Hjørland, Birger. 2004. A rejoinder to Beghtol (2004). *Knowledge organization* 31: 199-201.

Reynolds, Paul Davidson. 1971. *A primer in theory construction*. Indianapolis: Bobbs-Merrill.

Shera, Jesse H. 1965. *Libraries and the organization of knowledge*, ed. and with an introduction by D.J. Foskett. London: C. Lockwood.

Zerubavel, Eviatar. 1991. *The fine line: making distinctions in everyday life*. Chicago: Chicago University Press.