Complex Concepts into Basic Concepts

Rick Szostak
University of Alberta
Tory Building, Edmonton, AB, T6G 2H4, Canada
Rick.szostak@ualberta.ca
780-492-7645; FAX 780-492-3300

ABSTRACT

Interdisciplinary communication, and thus the rate of progress in scholarly understanding, would be greatly enhanced if scholars had access to a universal classification of documents or ideas not grounded in particular disciplines or cultures. Such a classification is feasible if it is possible to understand complex concepts as some combination of more basic concepts. The first part of the paper is theoretical. Philosophers are far from consensus in their view of concepts. Yet there appear to be five main types of concept theory in the philosophical literature. Each provides some support for the idea of breaking complex into basic concepts that can be understood across disciplines or cultures, but each has detractors. It will be argued that none of these criticisms represents a substantive obstacle to breaking complex concepts into basic concepts within information science. The second part is empirical. Can we take the subject entries in existing universal but discipline-based classifications, and break these into a set of more basic concepts that can be applied across disciplinary classes? The author performs this sort of analysis for Dewey classes 300 to 339.9. This analysis will serve to identify the sort of 'basic concepts' that would lie at the heart of a truly universal classification, and thus allow a pragmatic evaluation of whether these can be understood across disciplines and cultures. It is argued that there are two key types of basic concept: the things we study (individuals, rocks, trees), and the relationships that exist among these (talking, moving, paying).

Keywords

Concepts, classification, interdisciplinarity

INTRODUCTION

To what extent is it possible to define or classify concepts such that these can be understood in a similar way across disciplines and cultures? If this is possible, then a universal classification of documents or ideas not grounded in particular disciplines or cultures may be possible. If not, then information scientists may wish to focus on domain-specific classifications and translations across pairs of these. Many information scientists take the latter point of view. In particular, Hjørland (2009) has argued that a pragmatic philosophical perspective supports *only* domain analysis. I have disagreed with this conclusion in several places (Szostak 2008a, 2008b, 2010a, 2010b).

.

The crux of our disagreement, it turns out, rests on whether complex concepts can be broken into a set of basic concepts that can be understood similarly across groups. Hjorland would argue that breaking complex concepts into basic concepts which can then be understood similarly across disciplines reflects a 'rationalist' epistemology. [He would argue that all concepts can only be understood in terms of theories and thus a web of other complex concepts that will inevitably differ across communities.] Yet he would admit that a pragmatic epistemology does not exclude taking a rationalist approach as long as this works (see Dousa 2010 for a discussion of the implications of different types of pragmatism).

The disagreement is important precisely because interdisciplinary communication, and thus the rate of scholarly advance, would be encouraged by the development of a truly universal classification. Most obviously, such a classification would make it much easier for scholars to identify works in other disciplines or fields that address their interests, and then also to communicate their insights back to all relevant disciplines or fields. Scholarly advance often depends on juxtaposing previously unconnected insights (Davies 1989). Yet information science can aid communication in a quite different and perhaps even more important manner. Scholars disagree, and often do not understand why they disagree. If disagreements arise because different scholars construct complex concepts from different combinations of basic concepts, or imagine different sorts of influence within or between things (but do not appreciate these differences explicitly), classification in terms of basic concepts can enhance scholarly understanding. Given the potentially huge advantages of a universal classification grounded in basic concepts, great care should be taken before accepting a philosophical argument that this is impossible.

For the purposes of this paper, 'basic concepts' will be defined as 'concepts that can readily be ascribed similar meanings across disciplines or cultures.' Of course, the word 'basic' implies that they are readily comprehensible because of their simplicity, but this simplicity need not be part of the definition (and thus we need not worry about defining simplicity). 'Complex concepts,' alternatively, may be defined as 'concepts for which a shared understanding is possible only within particular disciplines or cultures.' The main theoretical claim of this paper is thus: 'Complex concepts, so defined, can be broken into basic concepts, so defined, most or all of the time, with a degree of shared understanding suitable for the purposes of information science.' This claim (and thus the definition of basic concepts employed here) will be justified both theoretically and empirically in what follows.

It is useful to provide some clarification. Basic concepts are not univocal: they do not need to be viewed in precisely the same way across groups or individuals. They need only to permit enough shared understanding for the purposes of classification. As will be argued below, information scientists need to appreciate when borrowing from philosophy that philosophers generally strive for precise distinctions: a concept is judged to be univocal or not. The sort of question of degree that lies at the heart of the classificationist's concern with concepts – is a concept too ambiguous for the purposes of classification? – is rarely addressed directly by philosophers. Though the present author is very appreciative of the field of natural semantic metalanguage (see below), and its search for 'semantic primitives' that carry the same meaning across all languages, it is not the intent here to hold all 'basic concepts' to that standard. This is important both theoretically – we can accept that ambiguity is universal and focus on whether the degree of ambiguity can be reduced to acceptable levels for the purposes of classification – and practically: the natural semantic metalanguage community has not yet identified nearly enough semantic primitives to form the basis of a classification system (see Goddard n.d.). It is also important to appreciate that we are not engaged in a reductive exercise: it is neither practicable nor desirable to reduce all of our understandings to the level of subatomic particles. Our desideratum is merely that broadly shared understandings are possible.

Different disciplines or cultures may treat a particular complex concept (say, democracy) quite differently. If the main theoretical conjecture of this paper is correct, these differences can then be identified in terms of more basic concepts. Members of one culture may for example think that severe limitations on who can vote or run for office are compatible with democracy, while members of another do not. At the level of these more basic concepts, the two groups may be able to understand what each other is talking about. When the former East Germany claimed to be democratic, others could have pointed out that it failed their definition of democracy due to the lack of free elections with multiple and self-selected candidates.

Hjørland (2010, 1079) would define democracy as "A society in which the opposition(s) to the ruling power is (are) free to organize and has (have) opportunities to communicate their ideas in public media" He argues that his definition is pragmatic because it speaks to "what kind of society should we aim at" and how to organize our theories and concepts to achieve this. He thus unnecessarily conflates 'is' and 'ought.' It would be much better to define (as above) democracy in terms of what it is, and then engage in intelligent conversation about whether and how it can deliver certain (ideally well-defined) outcomes. Likewise, while I agree that 'freedom of speech' and

'open media' are necessary for 'democracy' to function well, this point is best made by carefully defining these distinct concepts and outlining the relationships among them rather than attempting to define one in a way that subsumes the others. That is, we should as scholars carefully specify and justify causal arguments rather than bury assumptions regarding several causal relationships in our definition of a single concept. Hjørland's claim that concepts should be 'understood as socially negotiated meanings that should be identified by studying discourses rather than by studying individual users or a priori principles' (2009, 1530) can thus be seen to rest on a preference for unnecessarily complex and ambiguous concepts that inevitably muddle the scholarly conversation. Since such concepts are not uncommon, the classificationist must still cope with these. Though Hjørland's 'democracy' can be seen as a complex of complex concepts, the classificationist can (after perhaps opining that the scholarly enterprise would be better served by a less ambiguous definition) still proceed to identify the various components and relationships among these. Note that we can do a better job of classifying works about democracy if we distinguish works that discuss the role of freedom of the press from those that address whether certain characteristics of democracy yield certain societal outcomes. That is, the basic concepts that arguably comprise the complex concept will prove more useful in classifying real documents (especially those that are clear in their arguments) than the complex concept itself. This provides a further rationale for establishing that complex concepts -- even those that one might think are unnecessarily obscure -- can in fact be broken into basic concepts.

This paper thus explores the degree to which it is in fact possible to understand complex concepts as some combination of more basic concepts. The first part of the paper is theoretical. Philosophers are far from consensus in their view of concepts. Yet there appear to be five main types of concept theory in the philosophical literature. Each provides some support for the idea of breaking complex into basic concepts, but each has detractors. It will be argued that none of these criticisms represents a substantive obstacle to the use of this approach within information science. It should be stressed that the purpose here is not to develop a new theory of concepts: I will leave that task to philosophers. Rather, the purpose is to identify what sort of classification projects -- with particular emphasis on the possibility of a truly universal classification – might be justified with respect to *all* major concept theories in philosophy. I will argue that that is the appropriate stance for the information science community to take toward the philosophical literature.

The second part is empirical. To what extent can one take the subject entries in existing universal but discipline-based classifications, and break these into a set of more basic concepts that can be applied across

disciplinary classes? This sort of analysis will be performed for Dewey classes 300 to 339.9. This analysis will serve to identify the sort of 'basic concepts' that would lie at the heart of a truly universal classification (and thus allow a pragmatic evaluation of whether these can be understood across disciplines and cultures). Before the empirical investigation itself can be pursued, a brief intermediate section motivates the analysis and operationalizes the terminology.

What sort of concepts are 'basic concepts'? This paper will argue for two key types of basic concept: the things we study (individuals, groups, rocks, trees), and the relationships that exist among these (talking, moving, paying). [The possibility of a third type of adjectival/adverbial properties will also be raised briefly.] This further clarification of the nature of 'basic concepts' comprises the second major theoretical argument of this paper. To be clear, we are arguing that most/all things and relationships can at least potentially be ascribed similar meanings across disciplines or cultures, and as a result all complex concepts that can be understood as some combination of things and relationships can also be ascribed similar meanings across disciplines and cultures. The theoretical and empirical analysis below will support these arguments.

CONCEPT THEORY

"Research into the nature of concepts is ongoing, in both philosophy and psychology, and there is no general consensus in either field as to the preferred theory of concepts" (Earl 2007). Any attempt by information scientists to ground our work in concept theory must be guided primarily by an appreciation of the lack of consensus among both philosophers and psychologists. It is not for information scientists to presume that we can identify the 'best' philosophical argument, but rather to ensure that the classifications we develop are to the extent possible consonant with the widest range of philosophical and psychological theorizing (Stock 2010, 1953 makes a similar argument with respect to epistemologies). While individual information scientists may have their epistemological preferences, as a field we should view any argument in favor of or against any particular classification that is grounded in only one philosophical approach as seriously incomplete. Thus, information scientists need not have a 'practitioner' understanding of that literature but should strive to ensure that any classifications they might propose are not grounded in premises viewed with justified suspicion by any large group of philosophers or psychologists..

While information scientists should not attempt to adjudicate philosophical disputes (at least not within the pages of journals of information science) they should carefully evaluate whether particular philosophical arguments

result in a substantive objection to a particular classification. That is, we should be pragmatic in our borrowing from philosophy. For the purpose of this paper at least, the main question that information scientists need to ask is whether philosophers seek a degree of precision in defining concepts greater than information scientists have come to see as necessary from the user's point of view. Philosophers have trouble defining a term such as 'freedom' to their satisfaction. This fact has not deterred information scientists from attempting to guide users to works regarding freedom (though of course we would like to do so even better). We seek some set of descriptors that allows one group to understand what another means. Information scientists need to ask of any philosophical concern regarding precision what it means in practice for particular approaches to classification. In this regard, it is notable that philosophers naturally focus much of their effort on the more intractable challenges of definition. The information scientist should appreciate the possibility that many of the concepts that must be grappled with in classification may be far easier to cope with. This is one reason that this paper combines theory and empirics: one contribution that information scientists can make to philosophical debate is to identify the proportion of concepts that display certain degrees of intractability.

Philosophers have long noted the ambiguity of language, and postmodernism has taken these concerns to an extreme that would deny the very possibility of scholarly understanding (Alvesson, 2002, 54). These philosophical concerns have not caused information scientists to reject the task of classification as either impossible or undesirable. The pragmatic question is whether the degree of ambiguity can be reduced in particular circumstances enough to allow a particular classification to adequately serve the needs of users.

Philosophers appreciate an important distinction between 'concepts' and 'words.' Concepts may be less ambiguous than words. Information scientists have generally appreciated that it is concepts that must be classified in most cases (in the form of controlled terms or symbols standing for those concepts), though this recognition is often implicit rather than explicit.

"Just as thoughts are composed of more basic, word-sized concepts, so these word-sized concepts—known as *lexical concepts*—are generally thought to be composed of even more basic concepts" (Margolis and Laurence, 2006). Much of the philosophical literature over the centuries has in fact tried with little success to identify the precise constituents of such complex concepts as 'freedom' or 'democracy' or 'justice.' While philosophers have not always defined 'complex concept' and 'basic concept' in the way that we have in this paper (indeed, the very

definition of 'concept' is highly contested; Durbin 1988, 51), the fact that their purpose was to provide universally shared definitions of concepts otherwise viewed in different ways by different people implies that their approach was broadly similar.

The survey below differs from that pursued by Hjørland (2009) in several ways. First, it focuses on concept theories rather than the broader epistemologies in which these may be grounded. Since it is concept theory that has the most obvious implications for information science, a survey that is respectful of philosophical differences of opinion should focus its attention at that level. Philosophers have developed five broad approaches to trying to understand why in practice it has proved so difficult to realize the premise that complex concepts are composed of basic concepts (Margolis and Laurence, 2006). This survey does not attempt to argue for one best theory but rather seeks implications for information science in all five prominent concept theories. Since both surveys start from an appreciation of the diversity that exists across scholarly communities, it seems more appropriate to ground our classificatory efforts in the best insights from all concept theories than to choose only one. [Hjørland 2009 appreciates that the theories surveyed below represent a mainstream view of concept theory, though he worries that important but less prominent views are excluded; he nevertheless most closely associates his preferred approach with the 'theory theory' to be discussed below.] Most centrally, the survey that follows is oriented toward the one simple but central question: can (many/most) complex concepts be broken into basic concepts that can be understood across groups?

The Classical Theory

The classical theory argues that "a lexical concept *C* has definitional structure in that it is composed of simpler concepts that express necessary and sufficient conditions for falling under *C"* (Margolis and Laurence, 2006). This was important for philosophy, in that philosophers could then try to define terms such as 'freedom' in terms of more basic concepts, and once successful such definitions would allow a universally shared categorization of the free and unfree. The classical theory dominated the Western philosophical tradition's approach to concepts until Wittgenstein (1958) and even after (Earl 2007). It is still staunchly defended by many philosophers (including Jackson 1998 and Pitt 1999).

As noted above, the main objection has been empirical: though many philosophers through the centuries have attempted precise definitions of concepts such as 'freedom' they have collectively been much less successful

than they wished (Fodor 1981). Information scientists have much to contribute to the empirical side of this debate. Classification has proceeded nevertheless and with considerable success: both enumerative and faceted approaches assume some ability to break complex concepts into more basic. Yet this has tended to be done only within particular disciplines. Does this historical fact merely reflect the dominance of disciplines at the time that our major classifications were developed, or does it reflect difficulties in cross-disciplinary communication? [Gnoli (2006) explores how facets can work in non-disciplinary classifications.] This question can only be answered by exploring empirically whether most/all complex concepts can indeed be broken into basic concepts that can carry a shared meaning across groups.

The failure of philosophers to identify a set of necessary and sufficient conditions for precise identification of a complex concept has encouraged a 'neoclassical' theory (viewed by some as completely distinct from the classical theory) that stresses only the pursuit of necessary conditions (plus perhaps an incomplete set of sufficient conditions). The philosophical objection to the neoclassical position is that, in the absence of a complete set of sufficient conditions, it becomes possible for more than one distinct concept to have the same set of necessary conditions (see Earl 2007). But this is hardly a problem for information science. If information science will produce a hierarchical classification of concepts, where each concept appears in one unique place, then any set of subclasses will invoke one and only one broader class. [Classes and subclasses are the entry terms in classifications. Classes should ideally correspond to either complex or basic concepts, but in the former case the complex concept should be a combination of basic concepts. Subclasses are an ideally exhaustive set of 'types of' (or occasionally 'parts of') a class, such as 'poodles' are a subclass of 'dog.'] It is a standard principle when developing knowledge organization systems that when there are two synonymous or nearly synonymous concepts, only one is employed (Stock 2010, 1954). While information scientists should strive for an exhaustive listing of subclasses, the neoclassical position on concepts suggests that a nearly-exhaustive set of subclasses will provide a suitable clarification of the meaning of any concept. A listing of most but not quite all types of dog will adequately identify the nature of the 'dog' class.

There is a second problem with classical (or neoclassical) theory: typicality. People think apples are more typical fruit than are plums. If the concept 'fruit' lends itself to a simple definition, this should not be (Murphy 2002, Prinz 2002). Information scientists may see an opportunity where philosophers see a challenge: any cross-group differences in categorization of fruit may reflect differences in which characteristics of fruit are deemed most important. Moreover, note that classificationists must of necessity provide fairly exhaustive lists of subclasses: by

listing both plums and apples as subclasses of 'fruit' we declare that each deserves to be a member of that class. The 'problem' of typicality suggests that we explore some sort of intermediate subclasses which would appreciate that apples and plums are different kinds of fruit (though we should also appreciate that typicality may simply reflect commonality: people eat more apples than plums).

The Prototype Theory

All of the other theories are attempts to react to the empirical challenges to the classical theory. The prototype theory, a reflection of the problem of typicality noted above, and with roots both in Wittgenstein's philosophy of language and in psychology, suggests that instead of precise definitions we have probabilistic tendencies associated with concepts: something with characteristic A is more likely to be judged an X than something with characteristic B. Whether it is judged an X depends upon how well it compares to other members of the class X along many dimensions.

The implication for information science would be that we will always have difficulty defining classes and subclasses in a manner that all users will appreciate (though Stock, 2010, 1955, argues that appeals to prototypes provides a solution to fuzzy boundaries between classes). Users that think primarily of the characteristic 'color' when classifying fruit will seek plums elsewhere. Though prototype theory is a challenge to any classification, it *may* be less of a challenge to discipline-based or culture-based classifications if it is thought that members of these groups tend to stress the same characteristics when deciding where a subclass best belongs.

But prototype theory clearly is not the whole story, for people think logically about the nature of concepts, not just associatively. This is especially the case for a complex concept like "chairs bought on Wednesday" that can only be made sense of in terms of simpler concepts (Fodor 1998). Moreover, experiments find that individuals often identify '3' as an odd number more readily than '7,' even though both are clearly odd numbers (and indeed 'odd number' is the sort of concept that does lend itself to a classical definition: integers non-divisible by two). Thus, experimental findings of typicality effects need not imply that individuals do not define concepts logically (Armstrong et al, 1999).

Though rarely appreciated, it follows that an *exhaustive* classification of sub-classes does allow a very precise (albeit lengthy) definition of a class (Stock 2010, 1956, argues that classification is not just helpful but essential to precise definition). Information scientists often provide exhaustive or nearly exhaustive lists of sub-classes. They can thus provide an empirical rebuttal to the philosophic concern that the prototype approach is incompatible with the pursuit of clarity. Though particular users may think that plums belong somewhere else, they can be guided to appreciate that plums are 'fruit' within a classification that defines 'fruit' in a particular way (that happens to reflect the thinking of the vast majority of scholars of fruit).

Note also that if different disciplines or cultures differ regarding the members of a particular class, these differences can potentially be identified precisely and then understood across groups (as in the example of different definitions of democracy above). If some culture treats plums as a vegetable or a 'food despised by the gods,' the precise reason for this difference in classification can be provided. Users may still disagree regarding the 'best' classification, but can agree on the logic underlying different classifications. Notably, users will be able to successfully navigate a classification that they understand, even if they would prefer a different one in certain respects.

In sum, then, while the prototype theory provides warnings, it does not provide an unsolvable barrier to the project of a truly universal classification.

The Theory Theory

Theory theory has become the more popular approach in recent decades, and has had considerable influence on information science. It argues that "concepts stand in relation to one another in the same way as the terms of a scientific theory and that categorization is a process that strongly resembles scientific theorizing. It is generally assumed, as well, that the terms of a scientific theory are inter-defined so that a theoretical term's content is determined by its unique role in the theory in which it occurs" (Margolis and Laurence, 2006).

Thus even children when asked about whether a dog surgically altered to look like a raccoon is still a dog will draw on a simple biological theory to say yes (Gelman 2003). While the theory holds that people have an idea of the intrinsic nature of a concept (e.g. the essence of 'dog'), it suggests that individuals will disagree because each concept is understood within a web of concepts and theories. It had been hypothesized that children radically reorganize their concepts as they mature and develop richer theoretical understandings, but this seems not to be the

case. The empirical finding that concepts are stable over time has led some philosophers to wonder if they might also be reasonably stable across individuals and groups (see Spelke 1994). Certainly, information scientists might well suspect that the concept 'dog' is fairly well understood across disciplines and cultures (though the boundary with 'wolves' may need patrolling). More centrally, the dog example is one where the theory points at least potentially to defining characteristics (in this case genetic) which might be communicated across theories (evolutionists and creationists can both understand genes).

Hjørland and other information scientists draw on theory theory to suggest that concepts can only be understood within communities sharing a theoretical outlook. Note that the organizing principle of Hjørland (2009) is a post-Kuhnian view of the academy as comprised of competing paradigms that are only imperfectly understandable to each other. Since all concept theories are examined through a post-Kuhnian lens, the conclusion that concepts can only be understood from within a paradigm is not surprising. Such a guiding assumption virtually guarantees such a result. Philosophers worry, though, whether any two people would be able to understand each other if theory theory were correct. No two people share a complete set of theories of the world. If every concept can only be understood within an individual's complex web of theories and concepts, misunderstanding would be more common that it is (Fodor and Lepore 1992). Either concepts are understood in a manner that is not dependent on (anything approaching) one's entire web of belief's and/or individuals have access to strategies (such as hermeneutics) for understanding what another means without having to comprehend their entire web of beliefs. Either way, it must seem that theory theory cannot invalidate the possibility of shared understandings across quite different individuals (and thus groups) of at least some concepts.

That is, philosophers are aware of problems with both classical theory and its critics. Hjørland is confident that the critics will win. Indeed Hjørland (2009, 1520) assumes at the outset that "A shift in the understanding of concepts is part of a broader shift in our understanding of cognition, knowledge, and information." Yet it is of course dangerous to presume the future course in the history of ideas. Collins (1998) argues convincingly that skepticism of the possibility of scholarly consensus emerges when there are more than a handful of competing worldviews. That is, it is information overload rather than a direct experience of incommensurability that encourages skepticism of the possibility of cross-group understanding. And of course the best way of grappling with

information overload is via information science. We need not take the more extreme implications of theory theory as fact but can rather work to reduce the predicted level of ambiguity.

Hjørland (2009) quotes favorably Chalmers (1999) who argues that the concept 'mass' has clarity because the theory of Newtonian mechanics has clarity, whereas democracy lacks clarity because it is referenced in many ambiguous theories. Yet the argument can easily be reversed: theories of democracy are vague because they refer to vague concepts. While there *may* be a chicken-and-egg problem here, it is certainly easier to imagine clear concepts taking temporal precedence over clear theories: Newton thinking of a precise relationship among concepts that were already clear at least in his mind, rather than formulating the equation $F = ma$ and looking for concepts to fit. If we wish to encourage clear theories and concepts, then it makes sense to focus first on clarifying concepts. A clear theory of democracy requires clarity not only regarding democracy itself but (even more importantly) the influences it exerts on other phenomena (and vice versa) Clarifying concepts is certainly where information science has its greatest contribution to make.

Nevertheless, one practical implication of theory theory for information science is that it would be useful to classify scholarly documents in terms of the theory(ies) applied. Though beyond the scope of this paper, the desirability and feasibility of doing so is addressed in the León Manifesto (2007), Szostak (2004), and Gnoli and Szostak (2008).

Conceptual Atomism

Conceptual atomism argues that concepts are not related to each other but to the real world. Just as a personal name has no inherent meaning but rather refers to a particular thing in the world, concepts refer to real things in the world (Millikan 2000). Of course, no signifier ever reflects perfectly what it is supposed to signify (a recognition that forms the basis of the practice of semiotics and the pragmatic philosophy of Charles Sanders Peirce). And thus individuals, and perhaps especially those from different cultures, will likely perceive the world differently and derive different concepts from their perceptions (perhaps especially with respect to relationships). Nevertheless, we might hope that individual perceptions of the world are often similar enough that shared understanding of concepts is possible. Yet again we need to stress that we do not need to eliminate ambiguity but

merely reduce it to a degree that allows largely shared understandiungs of the concepts employed in a classification. If we all see enough dogs, and can communicate about what we see, we can agree (enough) on what a dog is.

Space prevents a detailed discussion of philosophical debate regarding 'reality.' Some postmodernists have noted that perceptions of reality differ, and have hypothesized that there is no external reality. Social constructivists have suggested that individuals construct their reality. Information science operates on the premise that there is an external reality that we can perceive, albeit imperfectly. Note that conceptual atomism would predict that differences in understanding would flow from differences in perception. While these in turn might be shaped by theory, it becomes much less clear why shared understandings would exist only within disciplines.

Conceptual atomism does not speak directly to the question of whether complex concepts can be broken into basic concepts. But it does make two critical arguments that are believed to be reasonable by many philosophers:

- Concepts do or should have external referents.
- An individual's understanding of a concept reflects their perception of the real world.

Taken together these arguments suggest that certain concepts – those that have clear referents in the real world – will lend themselves to shared understandings if individuals have similar perceptions. The question then arises of for what sort of concepts are individuals most likely to have shared perceptions of reality. The answer would seem to be those concepts that we observe clearly and repeatedly. Thus we should expect to have the greatest degree of shared perceptions of (many/most of) the 'things' that comprise the external world (dogs, trees). Somewhat less obviously, perhaps, we could also anticipate broadly shared perceptions of the relationships that exist among these things (walking, creating), for we do not just regularly observe static things but actions as well.

Notably, scholars of scientific ontology often argue that the best concepts to ground an ontology in are 'actually existing entities' (Merrill 2010, 83). While the most obvious type of basic concept is 'things that exist in the world,' our discussion of theory theory guides us also to classify the influences that these exert on each other. Happily, psychologists and anthropologists have found that relationships between things are perceived as 'real' by people, thus qualify as concepts, and are often perceived in similar ways across -- at least Western – cultures (Hudon 2001). And Vickery (2008) has noted that users of classifications may often want to search by type of

relationship. It will be argued below that a classification grounded in the things that comprise the world and the relationships that exist among these is feasible and supports understanding across disciplines and cultures.

Conceptual atomists – and especially Jerry Fodor – have made a strong claim that *all* concepts are understood innately through reference to the real world. Laurence and Margolis (1999, 62-3) argue that it is hard to imagine that this is the case for all concepts. They discuss how the concept 'father' might be built up from the concepts 'male and 'parent.' Though Laurence and Margolis question how any concepts might be established innately through reference to the real world, they nevertheless point to the possibility that some concepts may be combinations of others that are understood innately. Such an argument is obviously supportive of the two main theoretical conjectures of this paper: that complex concepts can be broken into basic concepts (from which they would have been built up in the first place), and that these basic concepts will in turn be concepts with clear external referents.

Conceptual atomism strongly encourages the identification of basic concepts as those which we perceive within an external reality. If people share an understanding (because of a shared perception) of 'dog' this is a basic concept. We need not dig deeper (though we could happily provide a reasonably exhaustive list of dog breeds, both to further clarify the meaning of 'dog' and because some users are interested in particular breeds). It is often noted that people best understand middle-level concepts such as 'dog' rather than either 'animal' or 'poodle.' Philosophers may wish to identify 'primitive concepts,' the most basic building blocks, but we are looking rather for common-sense shared understandings (Green 2002 makes a similar recommendation, both for practical reasons and from skepticism of the existence of truly 'primitive concepts').

Of course, our perceptions may themselves be informed by theory (as the dog example above illustrates). The ancient Greeks viewed water as a basic constituent of more complex substances whereas we now 'see' water as itself a combination of hydrogen and oxygen. Two points are called for: individuals can still share perceptions of the important characteristics of water (liquid, drinkable, and so on) without recourse to chemical theory; and the understanding of the chemical composition of water is itself widely shared. Nevertheless, it could be that some concepts (much more commonly in natural science than human science) that are not readily perceived (electrons, DNA), may be more theory-dependent than are rocks or dogs. Yet even here shared understanding seems quite feasible precisely because we can well imagine that these are 'things' in the world: sub-atomic particles are posited

to be components of all things, and DNA is posited to contain the basic information that guides the development of a range of life forms.

While conceptual atomists would stress that each basic concept is appreciated independently, it might nevertheless be argued that the concept 'dog' must be distinguished from 'wolf.' It might even be closely connected to the concept 'barking.' In practice, though, it does not seem that this minimal level of interconnectedness limits the ability of people to share broadly similar appreciations of 'dog' across cultures. Moreover, it cannot be stressed too much that the very task of classifying serves to further clarify the meaning of these concepts. If groups differ regarding whether 'foothills' are 'mountains' or whether 'marshes' are part of 'lakes' members of both groups should be able to understand a classification that clearly indicates whether these are included or not: is 'foothills' a subtype of 'mountains' or an alternative? Even in the seemingly more difficult case in which members of a group 'perceive' magma flowing from a volcano as the outpouring of some god's unhappiness, all can potentially agree on what magma looks and feels like, and thus their differences are not (at least primarily) about how to define concepts but are theoretical differences about the causal relationships at work. Members of the group in question can seek works on (unhappy)(god)(cause)(magma) or even (ritual)(cause)(god)(stop)(magma). The classificationist, then, should always seek to identify basic concepts such that cultural differences can be accurately seen as disagreements regarding causal relationships among shared concepts. This paper contends that this will often/usually be possible.

Though conceptual atomism can be critiqued as any other concept theory, information scientists should not ignore the implications of conceptual atomism: that concepts are (should be) linked to perceptions of the real world, and that individuals may perceive the things that comprise the world and the relationships among these in broadly similar ways.

Pluralism

Though philosophers prefer one compelling line of argument, there is increasing appreciation that each of the approaches above (and perhaps others) has merit. Information scientists should thus be wary of grounding their efforts in any one concept theory. While the classical idea that complex concepts can be broken into basic concepts (that lend themselves to shared understanding) is attractive (at least to this author), the information scientist needs to worry about whether this can be done in practice to the degree required by information science. While prototype

theory provides useful insight into how people make some classification decisions, the information scientist needs to appreciate that logic must play a role in such decisions, and also that an exhaustive or nearly exhaustive classification of subclasses can potentially provide shared understandings. While prototype theory thus warns us that members of different groups *may* perceive concepts differently, it does not establish that this degree of ambiguity is necessarily too great for the purposes of classification. Likewise, while theories undoubtedly influence our understanding of at least some concepts, this hardly means that they are entirely determinative, or that we cannot share theoretical understandings. Conceptual atomism guides us to ground concepts in aspects of the real world that we readily perceive (things and relationships); we must nevertheless be wary that some things and relationships are difficult to perceive. Conceptual atomism provides perhaps the most powerful rationale for the contention of classical theory that we can have shared understandings of (some) concepts across groups: our perceptions almost certainly differ less than our theories or beliefs (and thus we can imagine that we structure different theories and beliefs around broadly similar basic concepts). Though each theory has its critics, the best argument for not following any one of these is the fact that none have been entirely discredited.

Classical (and neoclassical) theory and conceptual atomism are each in their own way confident that concepts can be identified very precisely. Prototype theory is less optimistic, but suggests strategies that may support concept identification in at least some cases. Theory theory is the most pessimistic. The information scientist wants to know whether the degree of ambiguity predicted by concept theory is enough to render impossible certain classificatory projects. Classical (and neoclassical) theory, conceptual atomism (especially once amended rto allow complex concepts to be built up from basic concepts), and arguably prototype theory are supportive of the possibility of a truly universal classification. Theory theory argues that some degree of ambiguity is unavoidable. Yet as critics of theory theory have noted, this degree of ambiguity is clearly not enough to render everyday conversation impossible. Information scientists must be careful of leaping from a philosophical finding that 'there is ambiguity' to a conclusion that 'there is too much ambiguity.' Philosophers have been seeking flawless definitions. As noted above, information science does not need that degree of precision. We must accept, then, that philosophers will have devoted little attention to the 'how much?' question that must guide information scientists with respect to ambiguity. Especially once the important criticisms of theory theory are taken into account – and it is appreciated that other credible theories are *much* less pessimistic – it must remain an open question (to be returned to empirically

below) as to whether the degree of ambiguity posited by that theory is enough to interfere with the project of a universal classification.

Philosophers suspect that different sorts of concepts may best be understood in different ways. If so, this means that information science can make an important contribution to philosophical debate: by attempting universal classifications of concepts, we can identify which concepts seem to best fit the different views of concepts outlined above. Which concepts have clear external referents? Which concepts can be seen as a certain combination of concepts that have clear external referents? Which concepts are only comprehensible within a particular theory? The empirical section of this paper describes how this sort of analysis might proceed.

We can close by briefly reflecting on the broader implications of a pluralistic approach. If it makes sense to see different concept theories as complements rather than substitutes, the same may be true of theoretical differences in other spheres. If so, this would provide a further reason for expecting that concepts might be comprehensible across theories. Nevertheless, it bears repeating here that it would be very useful for users if documents were classified in terms of the theories and methods (and perhaps philosophical and disciplinary perspective) applied.

Some Thoughts On Implications for Information Science

The foregoing analysis of concept theory is broadly supportive of facet analysis, but arguably of a particular sort of facet analysis. While the present author considers himself as operating within the tradition of facet analysis, he has long worried that applications of facet analysis are unnecessarily unwieldy. The analysis here suggests that part of the problem may lie in the simple fact that 'facet' is not itself a basic concept in the sense that the concept is readily appreciated. The way forward involves clearly distinguishing facets from relationships (as Vickrey 2008 has advocated), stressing causal relationships, and treating the most common facets in away that is both manageable and readily understood across disciplines (sadly, most facet analysis proceeds within disciplines, with the notable exception of the freely faceted ILC). Of the thirteen facets identified within the Bliss classification, three -- thing, kind of thing, and part of thing – would each be captured within a hierarchical classification of 'things.' No facet indicator would be necessary if the classification made clear when the hierarchy was organized around types versus parts. Two other facets – product and byproduct – would also be captured in a classification of things, and their role as results of a causal process would be clearly indicated using linked notation of the sort

‘(A)(causes)(B).’ Likewise ‘agent’ would appear as the instigator of a causal relationship. ‘Patient,’ by which Bliss denotes ‘intermediate goods’ would be captured by ‘(A)(causes)(B)(causes)(C).’ Two important facets – time and place – are nevertheless simply particular types of things; the classification of things should devote particular attention to these (see Appendix 1), and linked notation can be employed to indicate time or place as necessary. Likewise ‘materials’ are things and can be handled with linked notation. Two facets – process and operation – are properly relationships (the first being internal), and can potentially be addressed through a universal classification of (especially causal) relationships. That leaves one facet – properties – that has not been addressed above. It is intriguing that philosophers have focused the vast bulk of their attention on things and (less so) relationships, and have said little about (adjective or adverb-like) properties. We might nevertheless conjecture here that certain properties at least – big, pretty, fast, easy – might also deserve the status of ‘basic concept’ in that individuals may share very similar appreciations of what they mean (note that we do not have to agree on what things are ‘pretty’ in order to share an understanding of the concept ‘pretty’). In sum, then, the analysis above seems to point toward a classification that embraces facets but eschews facet indicators in favor of a reliance on linked notation between universal classifications of things and relationships, and likely properties as well.

As such, the analysis is also supportive of post-coordinated indexing. Users should be able to search by the combinations of things and relationships (and properties) that are of particular importance to the user, rather than being forced to search only in terms of the compounds generated by the classificationist. Post-coordinated searches will be most efficacious if all complex concepts are first translated into basic concepts. If Hjørland’s ‘democracy’ is not unpacked into its constituent parts then searches for ‘freedom of the press’ will miss works identified only in terms of his complex ‘democracy’ amalgam.

Thesauri tend to focus on nouns, translate verbs into nouns (‘clean’ becomes ‘cleaning’), and ignore adverbs and adjectives. The analysis above suggests that verbs, adjectives, and adverbs deserve more attention. More centrally, the analysis above is a challenge to the common practice of generating thesauri only for particular fields of study. It should be feasible to pursue a universal thesaurus grounded in basic concepts.

Information scientists have become increasingly interested in formal ontologies in recent years. Ontologies involve creating a logical structure for language expression. Though the ontological project can be traced to at least Aristotle (who was much less accepting of ambiguity than Plato), the desire to allow computers to process spoken

language has been a key motivator of ontological research. Ontologies structure lexicons by including descriptive logics. Since each concept must be carefully defined in terms of a particular logical structure, creating ontologies is a slow, time-consuming and controversial practice (see Masolo et al, n.d.). The development of ontologies in the twenty-first century bears some resemblance to the development of classification systems over a century ago: many competing systems exist grounded in conflicting principles. While ontologies may hold great promise for information science (though most ontologies have been developed outside the field), it would thus be difficult at present to ground a universal classification in an ontology. The approach taken in this paper may be seen as a middle ground between present classifications and formal ontologies: it calls for adherence to logical rules in developing hierarchies of things and relationships, and argues that complex concepts can and should be understood as combinations of basic concepts, but does not demand that either basic or complex concepts be defined precisely in terms of some logic. Indeed, this paper has repeatedly accepted that some degree of ambiguity may be unavoidable – both in defining basic concepts and in defining complex concepts in terms of basic concepts – but is acceptable for the purposes of classification. It is thus possible to structure a classification grounded in basic concepts *right now*; it could be that as ontologies are developed and some consensus is achieved on which ontology is best suited to the needs of information science ontological insights will allow further clarification of both basic and complex concepts. In other words, the approach of this paper is consonant with that of ontologies, but by demanding less precision than ontologies it provides, at least for the foreseeable future, a more practical way of structuring a universal classification.

SETTING THE STAGE FOR AN EMPIRICAL INVESTIGATION

Our theoretical discussion can only raise some concerns for the classificationist. The question of whether a truly universal classification is feasible cannot be decided on theoretical grounds alone. This paper thus pursues an empirical test. In moving from theory to empirics there is generally an intermediate step in which theoretical terms are first operationalized, and the precise methodology to be employed is outlined. These introductory comments and the next two sections perform these functions. In particular, it is important to consider what sort of empirical analysis would best establish the feasibility (or not) of a universal classification.

One question raised above was whether the practice of organizing universal classifications by discipline was a historical accident or rather reflected real difficulties in developing shared understandings across disciplines. One obvious path for empirical investigation, then, is to take the concepts classified within a discipline-based universal classification, and see if these can be rendered in terms of a classification of basic concepts that potentially could be understood across disciplines.

Efforts to develop a new classification should always proceed using a mixture of deduction and induction. That is, the classificationist should first think logically about the broad structure of the classification, and deduce as far as possible the classes that fall within it. The classificationist can then flesh out the classification inductively by making sure that all entities that one wishes to classify have a place within the classification (it may at times prove valuable to ask if concepts that often co-occur belong in the same class or are causally related). If complex concepts can in practice be broken into basic concepts, this balance between deduction and induction is readily achieved: a logical division is pursued to the extent possible and supplemented with an examination of how concepts are employed in practice. If, however, complex concepts can only be comprehended within a web of concepts and theories, there may be little scope for deduction. [There *may* still be scope for deduction and the decomposition of complex into simple concepts *within* a given theory: the problem would consist in the mapping of these concepts across theories.]

Literary warrant often supplies a source for the inductive component of efforts at classification. For the purposes of this paper, the entries in an existing classification can serve the same purpose: can these be readily translated into a set of basic concepts? In order to answer this question the broad outlines of such a classification of basic concepts must first be sketched. It was suggested above that there are two main types of basic concept: the things we study and the relationships between these.

Basic Concepts 1: The Things We Study

Szostak (2004) provides a table of the main subjects of study in the social sciences and humanities (hereafter human science). The table has two individual-level categories (genetic predisposition and individual differences) seven societal categories (economy, politics, culture, social structure, population and health, technology and science, and art) and a non-human environment category. Each of these categories was subdivided and then generally subdivided again (some fifty different values are listed under cultural values). It was argued that the

phenomena in the table could be understood in terms of either their essence (most common for natural phenomena) or function (most common for social phenomena). This table can be fleshed out and extended to natural science through recourse to the Integrative Levels Classification (ILC; see ISKO Italia 2004).

The table in Szostak (2004) was not designed with the purpose of classifying scholarly documents or ideas, but rather for characterizing the broad outlines of the scholarly enterprise. Several elements in that table were thus subjected to greater efforts at logical subdivision (see Appendix 1). Most obviously the entries for time and location need to be treated in much greater detail. Usually, subdivision occurred in terms of 'types of': thus within the class of '[human] genetic predisposition' there are over twenty types of ability (including five types of perception, six types of decision-making, and six types of learning), ten motivations (food, sex – which is further disaggregated – shelter, achievement, etc.), and thirteen emotions (some of these could be further subdivided as necessary). In a minority of cases, though, subdivision instead occurred logically in terms of 'parts of': Each of the main dimensions of human personality subsumes more narrow characteristics: conscientiousness subsumes thoroughness, precision, foresight, organization, and so on. Note in this case that these narrower terms lend themselves to cross-group understanding much more than the broader term. If an exhaustive list of these can be generated, then understanding of the broader term is facilitated.

Note that other classifications often 'abuse' hierarchy: 'recycling ratio' is not a subclass of 'recycling' but is treated as if it were (Mazzocchi et al, 2007). 'Recycling rate' should instead be captured through linked notation; subdivision should proceed logically.

These logical efforts generate a schedule of some thirty pages in length (abridged in Appendix 1), containing several hundred entries (Szostak 2011a). This schedule is primarily focused on human science, though sufficient detail on natural science is provided to facilitate the classification of most works that connect human and natural science. This schedule would need to be supplemented by detailed lists of such things as countries, occupations, or industries. Such lists can generally be imported from other sources.

The result then is a very compact list of basic concepts. There can be little doubt that this table will expand in size as further inductive efforts are undertaken. The exercise that is described below, in which several hundred pages of Dewey Decimal Classification (DDC) classes were translated into combinations of basic concepts (see examples below), suggests that the degree of expansion required is in fact quite small. Nor should this result be

entirely surprising. If each complex concept were merely the combination of two basic concepts, then a list of 500 basic concepts could potentially generate some 125,000 complex concepts (500 times 499 divided by two, if one rules out 'A plus A' and does not wish to double-count 'A plus B' and 'B plus A') . As we shall see, many complex concepts involve more than two things; they also generally involve some relationship (see below).

The use of linked notation has not in the past resulted in such a dramatic decrease in the size of schedules needed for a library classification. But linked notation has been employed previously as a supplement to a system of classification in which complex concepts were not uniformly broken into constituent basic concepts. Classification at the level of basic concepts – if possible – promises a degree of simplicity in classification schedules heretofore unimagined. Universal Decimal Classification (UDC) uses linked notation extensively, and is often criticized for unwieldy notation. Much of the problem stems from very long notation for simple subjects. UDC may also sometimes put more detail in the call number than is required (Foskett 1996, 186). Though notation is beyond the scope of this paper, it should be appreciated that very brief notations can be given to each class in this schedule, and thus complex concepts which involve as many as five basic concepts can be handled with notation of manageable length.

Do the various entries in these schedules qualify as 'basic concepts'? That is, are they subject to common understandings across groups? This is, of course, a judgment call. In many cases, a positive judgment seems to be indicated: sight, food, fear, height, parent/child relationship, wages, telephone, sculpture, nuclear family, architect, farmer, teenager, job description, international organization, dictionary, mathematical model, fertility, and even attitude toward violence. In such cases, it should be possible to achieve a very high degree of agreement. Take 'food' for example: while people certainly disagree about which items they are themselves willing to eat, an understanding of food as something digested which is thought to provide energy to the body seems entirely capable of shared understanding. Certainly it would seem that such terms lend themselves to a much greater degree of shared understanding than complex terms such as 'freedom.' The claim of this paper is that the classificationist can be comfortable employing such concepts in a universal classification.

In many cases, the act of classification itself clarifies meaning. As noted above, personality dimensions become clearer when subdivided. The meaning of 'love' becomes clearer if 'love' is distinguished from 'joy' or 'jealousy' or 'guilt' or other emotions.

Nevertheless, it might be argued that some terms are not really basic concepts. For example, it is necessary to classify 'political ideologies' (since many works are written about these). Again one can clarify the meaning of 'socialism' by contrasting it with 'liberal democracy' and 'conservatism' Some important differences in understanding may still remain across both disciplines and countries. Notably, these might be reduced by still further efforts at subdivision. Socialism might be clarified by providing a list of types of socialism (as prototype theory above might suggest).

The manageable scope of the schedules is noteworthy in this respect. While one can readily despair of achieving shared understandings of the hundreds of thousands of classes in DDC, it is much more possible to aspire to shared understandings of some tens or hundreds of non-obvious basic concepts. This is especially the case given that these are placed within a logical hierarchical structure of manageable scope.

Basic Concepts 2: Relationships among Things

While all extant systems of document classification make some attempt to classify things (within disciplines), efforts to classify relationships are both less extensive and less successful. Despite the centrality of relationships to knowledge organization (and advancement), existing knowledge organization systems fail to express these as usefully as they could (Green 2008, 154). Indeed the knowledge organization literature is littered with suggestions for such a classification (see Perrault 1994 for a survey); none have ever achieved broad appeal. It behooves a classificationist to think carefully about how to render such a classification attractive.

The starting point for such an effort should involve reflection on what sorts of relationships will most often need to be referred to. A perusal of the scholarly literature establishes that most scholars study causal relationships. That is, they investigate *how one or more things influence one or more other things*. Some existing classification systems *allow* causal links to be coded by agent facets and/or by influence phase relationships (Ranganathan 1967, sect. SD) but none takes the coding of causal links as one of its major purposes. The León Manifesto (2007) urged a focus on relationships.

Complex concepts likewise generally refer (albeit often vaguely) to how some things influence others. 'Patriarchy' refers to how some set of cultural values and institutions constrains various choices available to men and especially women. 'Globalization' refers to how closer economic, political, and/or cultural ties affect domestic

economies, polities, and/or cultures. If we are to classify scholarly literature, and cope with complex concepts, then we need most of all to classify 'types of influence/causation.'

Two important caveats immediately arise. First, as the examples in the previous paragraph suggest, scholars are often not clear about what type of influence is being discussed. Many works and concepts, then, must be described at the general level of 'X influences Y.' Second, there are of course non-causal relationships that must be appreciated. 'X and Y,' 'X from the perspective of Y,' 'X for Y,' and 'X compared to Y' are prominent examples. Note, though, that in principle we need not worry about 'X as an example of Y' or 'X as an element of Y,' for these relationships are already captured within the structure of the classification of things. In practice, though, it is still useful to use linked notation for a complex concept such as 'map library' rather than provide an exhaustive list of all types of library. The above examples exemplify the six broad types of relations identified by Perrault (1994). It is notable that Perrault treats causation/influence as just one of six types of relationship, and devotes less effort to subdividing this than other types of relation.

What types of influence are most important? There is no widely accepted logical classification of types of influence in the literature. Thousands of different causal verbs have been identified; Khoo (1995) organized these into 47 broad types. Note that the absence of a well-developed classification of causal relationships biases our empirical test against the finding that a universal classification is feasible: such a classification would act (as does the classification of things) to clarify the meaning of many terms. Yet the test can proceed nevertheless, with the investigator identifying the relators that seem necessary to translate complex DDC entries into basic concepts, and then asking whether these relators qualify as basic concepts.

Though the development of a classification of causal relators is a work in progress (Szostak 2011c), the empirical test performed for this paper found that the vast majority of DDC entries confronted could be handled through recourse to a mere handful of terms:

- moving (using physical force)
- controlling/ supervising
- paying/financing
- talking to
- selecting from

- evaluating, judging
- preventing
- conflicting [a broad term capturing any sort of disagreement; it is often useful to identify particular types of conflict]

It is argued that these each qualify as basic concepts: people across disciplines and cultures can agree on what these terms mean (though members of barter economies might grapple with 'paying'). These basic relationships can be used in combination. Controlling plus paying might refer to providing financial incentives (with bribery as a special case involving the breaking of rules). The non-existence of these relationships may at times be important: not controlling, not paying, and so on. By allowing combinations and negatives, even this small list of basic influences generates a much larger set of possible relationships.

The above list of verbs captures (for the most part) causal links between phenomena. For works that focus on (at least in part) changes within a phenomenon, the following verbs are by far the most commonly invoked (and could be considered to be 'internal relationships'):

- grow or develop
- decline
- fluctuate or cycle
- exhibit stability

These can be combined with an indication of causation to generate 'maintains,' 'causes to increase,' and so on.

It makes sense to rely somewhat more heavily on induction in developing a classification of relationships than was the case with things. If key terms are provided with simple notation, a wider set of terms can be utilized more sparingly.

AN EMPIRICAL INVESTIGATION: TRANSLATING DDC CLASSES 300 TO 339.9

Is it possible to translate the classes of a major existing universal discipline-based classification into the basic concepts outlined above? If so, how many basic concepts are required to identify the complex concepts identified in the classification? And which basic concepts prove most important? These questions were explored with respect to class numbers 300 through 339.9 in DDC (see Szostak 2011b).

The broad results of that test can be briefly summarized. The author found it straightforward to translate almost all DDC entries. When it was not obvious how to proceed, this was usually the result of ambiguity in DDC terminology. In many cases, the DDC heading itself was a combination of two or three basic concepts [Yet within DDC it is not generally straightforward to search by the basic concepts within such compounds.]. As an example, DDC 302.35 'Social interaction in complex groups' can be translated as (Interpersonal relationships) (in) (groups) [It is noteworthy here that the subdivisions in DDC refer to organizations rather than groups. The schedule of basic concepts lists types of both groups and organizations. In this case as many others, translating into basic concepts serves to clarify DDC terminology.] In all cases, a manageable handful (or fewer) of basic concepts was all that was required, and thus these could potentially be represented notationally in a quite manageable notational space.

How, though, to establish within the limited space of a paper that these results do not reflect authorial bias? It proved tedious to provide a random sample of the entries in Szostak (2011b) (the author experimented with providing every twenty-fifth entry), for most entries were unproblematic and thus of little interest. The approach to be followed here, then, is to discuss at some length some of the more challenging entries (within the subsample of every twenty-fifth entry). While such an approach cannot be a perfect check on authorial bias, it does provide the reader with an understanding of how the translation process proceeded and thus a sound basis for judging whether a universal classification in terms of basic concepts is feasible.

We can start with 303.376 'Censorship.' This is the sort of complex concept that one might think lends itself to quite different interpretations. Yet the essence of censorship is surely something like: (preventing) (publication). Arguably, each of these two components can be seen as basic concepts. They surely are less subject to ambiguity than the concept of censorship itself. And further clarity can be provided with respect to particular instances, by coding for the purpose of censorship: political, religious, cultural values, and so on. I confess that I first considered (government)(controlling)(publication), but it was pointed out that censorship is not always an act of government (though it will generally be useful to indicate who is doing the censoring) and that 'preventing' is more accurate than 'controlling.' There is still a slight problem in that censhorship is sometimes applied to performances; one could potentially substitute 'expression' for 'publication' to cope with this, but with some loss in clarity. It may be preferable to note in scope notes that censorship can be applied to performance, and then capture such works with the somewhat awkward (preventing)(publication)(performance). In any case, for present purposes the point is that it

is possible to render 'censorship' in terms of basic concepts, and scholarly conversation can serve to identify the optimal way of doing so.

Notably, differences of opinion regarding particular examples of censorship often reflect disagreements over whether a certain type of censorship is 'good' or 'bad,' rather than about whether it meets the sort of definition of censorship provided here. This is one reason why this concept (and many many others) can be rendered in terms of basic concepts fairly easily. Yet this observation guides us to want users to be able to find both descriptions of censorship and ethical evaluations of censorship when they search, *but to be able to readily distinguish works that are primarily descriptive from those that are primarily evaluative.* This can be done using basic concepts by adding (ethical evaluation)(of) to any heading – or if possible a more precise indicator of a particular type of ethical evaluation, such as applying theories of rights.

The twenty-fifth next entry is 303.69 'Conflict resolution.' This can be rendered as (Cause) (Conflict) (to end). I at first translated this as (conflict) (decline), but this conflicted with my translation of the main DDC subdivision of 'peacemaking' as (cause) (to end) (state) (violent conflict with) (state). [Note that notationally the last three terms would be grouped together to indicate that it is not states being ended but conflict between these.] The lesson here would seem to be that care must be taken in translation, *but that logical rules can be followed that generate one precise translation.* There is a problem with the rendering of 'peacemaking' for it refers only to conflicts between states. Most contemporary peacemaking is focused on conflicts within states. This problem can be solved by translating 'peacemaking' as (cause) (to end) (violent conflict), and then distinguish conflict between states from conflict within states.

We can turn next to 323.5 'Political rights.' As has already been noted, a classification of basic concepts must have entries for different types of ethical evaluation. Political rights are thus properly the application of this type of ethical analysis to matters of government. Particular rights such as 'right to assembly' can be precisely designated as (right analysis) (of type) (group of individuals) (speak to) (state). 'Rights' is, of course, precisely the sort of term that semioticians (among others) would suspect is interpreted differently across individuals. It is certainly one of the more ambiguous terms to be considered a basic concept here. It is argued nevertheless that, though contested, the concept of 'rights' allows a fair degree of shared understanding across groups. If this is not the case (say, if a survey of diverse users found radically different interpretations) , then it might be both possible and

desirable to render the concept ‘rights’ itself in terms of even more basic concepts (individuals/groups should be allowed/able to …), though the classificationist would then have to choose among definitions offered by different philosophers. Note that users are much more likely to disagree over an evaluative issue regarding rights (such as does a fetus have a ‘right to life’?) than about what it means to say ‘right to life.’ And thus users with quite different opinions on the evaluative question could navigate without difficulty a classification that guided them to both evaluative works and descriptive works regarding rights in general or particular rights (and it might be desirable to distinguish arguments in favor of a certain right – or even rights in general – from arguments against).

It is useful also to look at 332.41 ‘Value of money.’ Such an entry may at first glance appear to be an obvious compound of simpler terms. Yet in fact this vague term seems in context to mean “inflation/deflation”; and thus is best captured by a more precise (increase/decrease) (price level). Philosophical considerations of economic value are captured elsewhere in DDC and would be translated by linking ethical analysis to economic elements. This example provides further evidence that translation into basic concepts can serve to clarify the meaning of terms in existing classifications. And if such translation became automatic, then concepts added in future would be clarified at the outset.

What of 333.9 ‘Other resources’? It is hard to imagine a user searching by this heading. This diverse class would in some cases refer to subclasses of (mineral or energy) natural resources (which should all be listed in the schedules) and in some cases require linked notation. For example, biological resources could be (natural resources) (of type) (flora/fauna). Translation into basic concepts here allows much greater specificity and thus greatly enhances retrieval. The greater specificity is possible because of the reliance on linked notation among basic concepts.

The DDC entry 338.001 ‘Philosophy and theory of production’ is a typical catch-all class when a new subject is first encountered in the schedules. As noted above, different types of philosophical analysis must be identified as basic concepts, and then it is straightforward to identify the application of any/all of these to any topic. It is also possible to identify different types of ‘scientific theory’ (where scientific theories are defined as theories about how the world works). Note that it is highly desirable to distinguish scientific treatments of production from philosophical treatments (Is production good?; How should we study it?; and so on). In this case, most users will be

interested in either economic analysis of how and why production occurs or in various types of philosophical evaluation of production, but only the rare user will want both.

Discussion of Results

It is quite easy to translate the vast majority of DDC terms into the (very manageable schedules of) basic concepts. At present, DDC often provides entries under different disciplines for the same compound term. This need not be done if classification occurs in terms of basic concepts (works and even ideas could still be coded for discipline if desired). Moreover, searches by basic concept are often difficult or impossible within DDC, but would be straightforward after translation into basic concepts.

The translations provided above and in Szostak (2011b) were largely the work of the author. Further advances in clarity and consistency, and thus in reducing ambiguity, can be anticipated if a wider group of scholars were to be involved in such an exercise.

In the theoretical section of this paper, it was argued that information scientists should not seek an unattainable perfection in clarity of definition in the terms used in a classification, but rather seek enough clarity to provide useful classifications to users. DDC is far from perfect but is clearly useful. Translating DDC terminology into basic concepts is possible and would make the classification more useful in important ways. In particular it makes complex terms more understandable across disciplines.

While the translation can be conceived as a supplement to DDC (or other extant systems) it would yield even greater benefits over time if as new entries were added or old entries reviewed classificationists reflected on the basic concepts involved. Moreover, if classifiers were guided to reflect on basic concepts as they looked at particular works, they would often be guided to suggest novel combinations. Most scholarly works discuss how one or more things affect (often in particular ways) one or more other things: such works can be better classified if classifiers think in terms of compounds of basic concepts. If a classifier confronts a book discussing the effect of attitudes toward punctuality on employment patterns in country X, both their task and user retrieval are best served by the use of linked notation rather than for both to try to imagine a unique class heading.

This paper has been guided by a distinction between complex and basic concepts. For purposes of evaluation, there is value in thinking of ‘degrees of communicability’ or perhaps ‘degrees of non-ambiguity.’ That

is, some concepts can be more basic than others. In this respect, most DDC entries can be translated into 'very basic' concepts for which the degree of ambiguity is minimal. All DDC entries can be translated into concepts that are 'somewhat basic' in the sense that the degree of ambiguity, at least potentially, might be considered acceptable from the point of view of information science. They are certainly much 'more basic' than the complex terms translated. Recall that the translation of DDC entries often served to clarify the meaning of DDC entries themselves. Such clarification would be even greater if new DDC entries were made with translation to/from basic concepts in mind. The translation thus has the potential to enhance cross-disciplinary communication while also enhancing within-discipline communication. This *must be* the case if translation to basic concepts is a complement to an existing discipline-based classification such as DDC. The question of whether a stand-alone universal basic concepts classification can be superior on both counts is harder to judge: the result might well depend on how well thesauri were constructed to guide users from disciplinary terminology to basic concepts. If all DDC entries that reflect disciplinary concepts were included in such a thesaurus, then again improvements on both counts could be anticipated.

If so, then the use of basic concepts achieves a truly remarkable result: greater clarity in classification (and thus much better retrieval) through the use of linked notation among elements in much more compact schedules. Note that users will have better retrieval results whether they are searching within disciplines (since even within disciplines terminology is clarified) or (especially) across. Perhaps even more importantly, they will achieve better results whether they know precisely what they are looking for (for they can specify a precise causal relationship) or are performing an exploratory search (for they will find works across all disciplines that address a particular topic). In the latter regard, a universal classification in terms of basic concepts has great potential for enhancing the discovery of 'undiscovered public knowledge': connections between existing but dispersed pieces of scholarly understanding that are of critical importance to the advance of scholarly understanding (Davies 1989). It may prove that coding works by relationship is especially valuable in this context. As for the compactness of schedules it is worth noting that present classifications often provide instructions on how to code for time or place or individual in multiple places whereas a basic concepts classification would by its very nature do so only once. For this reason and others, both classifier and user could much more readily master a basic concepts classification than any existing classification.

A classification that would be easier and more efficacious for classificationist, classifier, and user is worth pursuing. The results of this paper suggest that it is feasible.

CONCLUSION

The research undertaken here suggests that complex concepts can indeed be broken into basic concepts. Basic concepts are defined as those which lend themselves to an acceptable degree of cross-group understanding for the purposes of classifying scholarly documents or ideas. It is argued that there are two broad types of basic concepts: the things we study and the relationships that exist among these. A third type – adjectival/adverbial 'properties' of a thing – might also be useful.

The theoretical section warned of the dangers of assuming that a philosophical finding of ambiguity (still less a philosophical assumption of ambiguity) necessarily translates into a finding of *too much* ambiguity for a particular classificatory project. It urged us to respect *all* prominent concept theories, and found that most concept theories provided strong support for the idea of a universal classification grounded in basic concepts, while none established that there was too much ambiguity for such a project to be feasible. While information scientists were urged not to try to solve philosophical disputes, the information science literature can potentially provide important empirical insight into how important particular philosophical concerns are in practice.

It may of course be argued that the complex concepts grappled with in the empirical section were not that complex. But if so that also would be instructive: it must then seem that the bulk of the literature we hope to classify is not in fact written in terms of those concepts that have most troubled philosophers through the ages. Nor should this be surprising, for most scholarly and general literature addresses the world as it is – and thus at least potentially involves concepts with obvious external referents.

DDC has general classes for "Philosophy and theory of X." It was suggested above that philosophical theory be distinguished from scientific theory, and where possible the particular type of theory be noted. This strategy provides a way of dealing with those terms such as freedom that have defied precise definition: users can be alerted to different treatments of these.

Clarifying the meaning of complex concepts is useful both within and far beyond the boundaries of information science. For classificationists, it makes it possible to pursue a universal, non-discipline-based

classification. For classifiers, it facilitates the accurate classification of complex works. For users, such a classification would allow searches by thing, relationship, and causal argument (things plus relationship). The first of these is complicated within existing classifications by the fact that the same thing may go by different terms in different disciplines, or be subsumed within quite different complex concepts. The second and third are barely possible if at all, since most relationships are not coded for (and certainly not consistently across a classification), and complex concepts are not broken into constituent parts. Yet these are the sorts of searches that (especially interdisciplinary) scholars wish to pursue. Such a classification would thus be a major boon to the scholarly enterprise, and to all societal activities that benefit from enhanced scholarly understanding. And this may not even be the most important benefit of classifying with respect to basic concepts. Once scholars from different disciplines have found each other, they need to understand what each other says. A classification that clarifies the meaning of complex concepts in terms of basic concepts (as defined in this paper) will reduce substantially the degree of interdisciplinary misunderstanding, and thus further enhance the rate of scholarly advance.

Though beyond the scope of this paper, it is worth noting that the manageability of a basic concept classification may provide an answer to an emerging question in information science: how to allow users to search simultaneously across multiple databases including traditional library classifications, archives and museum inventories, and the variety of information sources now available on the web. It is likely not reasonable to expect the latter databases to be translated into the terminology of an existing library classification. It may be feasible to imagine all being translated/translatable into basic concepts.

Of course, the development of a new system of classifying documents or ideas – whether as a complement or substitute for existing classifications – is not a venture to be entered lightly. The present paper suggests not only that it is feasible but that the schedules may prove surprisingly manageable in scope. While domain analysis can play an important inductive role in fleshing out these schedules, this paper has shown that the premise that information scientists should *only* pursue domain analysis is misguided. Classificationists should work toward the development of a universal classification of basic concepts.

ACKNOWLEDGMENTS

I thank Claudio Gnoli and Tom Dousa, and anonymous referees, for very useful advice.

REFERENCES

Alvesson, Mats (2002) *Postmodernism and Social Research.* Buckingham: Open University Press.

Armstrong, S. L., Gleitman, L. R., and Gleitman, H. (1999) What Some Concepts Might Not Be. In Margolis, Eric and Laurence, Stephen (Eds.). *Concepts: Core Readings*. M.I.T. Press. 225-259.

Chalmers, A.F. (1999) *What is this Thing called Science?* 3rd. ed. Brisbane: Queensland University Press.

Collins, R. (1998) *The Sociology of Philosophies: A Global Theory of Intellectual Change.* Cambridge: Belknap (Harvard University) Press.

Davies, Ron (1989) "The Creation of New Knowledge by Information Retrieval and Classification" *Journal of Documentation* 45:4, 273-301.

Dousa, T. (2010) Classical pragmatism and its varieties: on a pluriform metatheoretical perspective for knowledge organization. *Knowledge Organization 37:1*, 65-71.

Durbin, P.T. (1988) *Dictionary of Concepts in the Philosophy of Science.* New York: Greenwood Press.

Earl, D. (2007) Concepts. *Internet Encyclopedia of Philosophy.* http://www.iep.utm.edu/concepts/ Accessed May 16, 2011.

Fodor, J. (1981). The Present Status of the Innateness Controversy. In his *Representations: Philosophical Essays on the Foundations of Cognitive Science*. Cambridge, MA: MIT Press.

Fodor, J. (1998). *Concepts: Where Cognitive Science Went Wrong*. New York: Oxford University Press.

Fodor, J. and Lepore, E. (1992). *Holism: A Shopper's Guide*. Oxford: Blackwell.

Foskett, A.C. (1996) *The Subject Approach to Information.* 5th edition. London: Library Association Publishing.

Gelman, S. (2003). *The Essential Child: Origins of Essentialism in Everyday Thought*. Oxford: Oxford University Press.

Gnoli, C. (2006) The meaning of facets in nondisciplinary classifications. In G. Budin, C. Swertz and K. Mitgutsch (eds.) *Knowledge organization for a global learning society, Proceedings of the 9th International ISKO Conference*. P 11-18 Würzburg: Ergon, p. 11-18.

Gnoli, C. & Szostak, R. (2008) Classifying by phenomena, theories, and methods: examples with focused social science theories. In C. Arsenault and J. Tennis (Eds.) *Culture and Identity in Knowledge Organization, Proceedings of the 10th International ISKO Conference*, Würzburg: Ergon, p. 205-211.

Goddard, C. (n.d.) Natural Semantic Metalanguage. Website

http://www.une.edu.au/bcss/linguistics/nsm/semantics-in-brief.php Accessed Feb 10, 2011.

Green, R. (2002) Internally structured conceptual models in cognitive semantics. In R. Green, C. Bean and S. Myaeng (Eds.) *The Semantics of Relationships.* Dordrecht: Kluwer, 73-90

Green, R. (2008) Relationships in knowledge organization. *Knowledge Organization 35:2/3*, 150-9.

Hjørland, B. (2009). Concept theory. *Journal of the American Society for Information Science and Technology, 60: 8*, 1519-36.

Hudon, M. (2001) Relationships in multilingual thesauri. In C. Bean and R. Green (Eds.) *Relationships in the Organization of Knowledge.* Dordrecht: Kluwer, 67-80.

ISKO Italia (2004-) *Integrative Levels Classification*. Retrieved May 15, 2010 from www.iskoi.org/ilc.

Jackson, F. (1998) *From Metaphysics to Ethics: A Defence of Conceptual Analysis*. Oxford: Clarendon Press.

Khoo, C. (1995) Automatic identification of causal relations in text and their use for improving precision in information retrieval. PhD dissertation, Syracuse University.

Laurence, S., and Margolis, E. (1999) "Concepts and Cognitive Science." In Margolis, Eric and Laurence, Stephen (Eds.). *Concepts: Core Readings.* M.I.T. Press.

León Manifesto. (2007) Retrieved May 15, 2010 from www.iskoi.org/ilc/leon.htm

Margolis, E. and Laurence, S. (2006) Concepts. *Stanford Encyclopedia of Philosophy*. Retrieved May 15, 2010 from http://plato.stanford.edu/entries/concepts/

Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A. (n.d.) *Ontology Library*. Laboratory For Applied Ontology - ISTC-CNR. http://wonderweb.semanticweb.org/deliverables/documents/D18.pdf

Mazzocchi, F., Tiberi, M., De Santis, B., & Plini, P. (2007) Relational semantics in thesauri: some remarks at theoretical and practical levels. *Knowledge Organization 34:4,* 197-214.

Merrill, G.H. (2010) Ontological realism: Methodology or misdirection? *Applied Ontology 5,* 75-108.

Millikan, R. (2000). *On Clear and Confused Ideas*. Cambridge: Cambridge University Press.

Murphy, G. (2002). *The Big Book of Concepts*. Cambridge, MA: MIT Press.

Perrault, J. (1994) Categories and relators: A new schema. *Knowledge Organization 21:4*, 189-98.

Pitt, David. 1999. "In Defense of Definitions." *Philosophical Psychology* 12 (2), 139-156.

Prinz, J. J. (2002). *Furnishing the Mind: Concepts and Their Perceptual Basis*. Cambridge: M.I.T. Press.

Ranganathan, S.R. (1967) *Prolegomena to Library Classification*, 3rd ed. Bangalore: SRELS. http://dlist.sir.arizona.edu/1151/

Spelke, E. (2003). What Makes Us Smart? Core Knowledge and Natural Language. In D. Gentner & S. Goldin-Meadow (2003). *Language in Mind*. Cambridge, MA: MIT Press, pp 277-311.

Stock, W.G. (2010) Concepts and Semantic relations in Information Science. *Journal of the American Society for Information Science & Technology 61:10*, 1951-69.

Szostak, Rick (2011a) *Basic Concepts Classification*. http://www.uofaweb. ualberta.ca/economics2/pdfs/Szostak-Basic.pdf Accessed May 16, 2011.

Szostak, R. (2011b) *Translation table: DDC to Basic Concepts Classification.* http://www.uofaweb.ualberta.ca/economics2/pdfs/Szostak-DDC-Basic.pdf. Accessed May 16, 2011.

Szostak, R. (2011c) Classifying Relationships. Unpublished research paper.

Szostak, R. (2010a) Universal and domain-specific classifications from an interdisciplinary perspective. In C. Gnoli and F. Mazzocchi (Eds.) *Paradigms and Conceptual Systems in Knowledge Organization: Proceedings of the 2010 conference of the International Society for Knowledge Organization.* Würzburg: Ergon.

Szostak, R. (2010b) Comment on Hjørland's concept theory. *Journal of the American Society for Information Science & Technology 61:5*, 1076-7.

Szostak, R. (2008a). Classification, interdisciplinarity, and the study of science. *Journal of Documentation 64:3*, 319-32.

Szostak, R. (2008b) Interdisciplinarity and classification: a reply to Hjørland. *Journal of Documentation 64:4.*

Szostak, R. (2004) *Classifying Science: Phenomena, Data, Theory, Method, Practice*. Dordrecht: Springer.

Vickery, B. (2008) Faceted classification for the web. *Axiomathes 18,* 145-60.

Wittgenstein, L. (1958) *Philosophical Investigations*. 3rd Ed. New York: MacMillan.

Appendix 1: OUTLINE OF THE CLASSIFICATION OF THINGS IN THE BASIC CONCEPTS CLASSIFICATION [In progress]

The main classes treated below are (in order):

1. Genetic Predisposition
2. Individual Differences
3. Economy
4. Art
5. Social Structure
6. Politics
7. Technology and Science
8. Health and Population
9. Culture
10. Non-human Environment

Following the ILC, it is expected that the following main classes will be developed/borrowed over time: Particles, atoms, molecules, celestial objects, rocks, biological entities

1. **Genetic Predisposition** [subdivided by 'types of' throughout, except anatomy: 'parts of']

[Note that while this category was developed primarily with humans in mind, relevant terms can be connected to other species using linked notation (e.g. bird sight)]

Abilities: Consciousness [subconsciousness]; Vocalization; Perception [five types]; Decision-making [six types]; Tool-making; Learning [reading, writing, remembering, etc.]; Other physical attributes [examples of, including locomotion, eating]

Motivations: Food/drink [List here carbohydrate, vitamin, protein (enzyme)?]; Clothing; Shelter [Sleep?]; Safety; Sex [Many types]; Betterment; Aggression; Altruism or fairness; Identification with group; Play

Emotions: Emotional display; Love; Anger or Annoyance; Fear; Jealousy; Guilt; Anxiety; Humor; Joy; Fatigue; Empathy; Grief; Disgust; Aesthetic sense; Pain; Pride; Resentment or Bitterness; Lust

Time Preference (That is, how future and present are compared.)

Anatomy: Embryo; Organs; Tissues; etc.

2. **Individual Differences**

Physical Abilities: Speed; Strength; Endurance

Physical Appearance: Height; Weight; Symmetry; Shape

Energy Level: Physical; Mental

Intelligences: Musical; Spatial; Mathematical; Verbal; Kinesthetic; Interpersonal

Personality: [subdivided by 'elements of']: Sociability (Extro/introversion) [Talkative; Assertive; Adventurous; Enthusiastic (vs.) Reserved; Withdrawn; Anti-social]; Emotionality (Stable/moody) [Contentment; Composure (vs.) Anxiety; Self-pity]; Conscientiousness [eight subtraits]; Affection (Selfish/agreeable) [seven subtraits]; Intellectual Orientation (Holistic/analytical) [five subtraits]; Other dimensions [eight listed]

Disorders [types of]: Schizophrenia; Psychoticism; Compulsive [More to be added.]

Sexual Orientation [types of]: Asexuality; Heterosexuality; Bisexuality; Homosexuality; Transsexuality

Schemas: [subdivided by 'types of'] View of self; View of others; Causal relationships
Interpersonal Relationships [types of]: Parent/child; Sibling; Employee/r; Romance; Friendship; Casual; Sexual

3. Economy [elements of]

Total Output [types of]: (types of expenditure): Consumption; Investment; Government spending; Exports and imports; (types of income): Wages; Profits; Interest; Rent; Output of individual goods and services within individual markets [Could use SIC codes or NAICS codes]

Economic Fluctuations: Price level; Unemployment

Income Distribution [types of]

Economic Ideology [types of]

Economic Institutions [types of]: Ownership [15 types]; Production [several types]; Exchange [Many types]; Trade [several types]; Finance [many types]; Labor relations [several types]; Marketing [several types]; Organizations [several types]

4. Art [types of]

Non-reproducible: Painting [Subdivisions can be further subdivided.]; Sculpture; Architecture; Prose; Poetry (rhyming or not); Collage; Cartoon; Graphic art

Reproducible: Theater; Film; Photography; Music; Dance

5. Social Structure [elements of]

Communities [types of]: Subdivided by scale: International; National; Regional; State/province; Local; Informal groups; Networks

Social role in general. [But specific roles should be specified when possible.]

Gender [types of]: Female; Male; Other

Family Types [types of]: Nuclear; Extended; Single parent; Kinship systems [can be subdivided.]

Marital status: Single; Engaged; Married, monogamous; Married, multiple; Common law; Divorced; Widowed; Separated

Classes (various typologies): Castes, caste system; Upper class; Middle class; Working Class; Underclass/ Disadvantaged

Occupations (various): Professionals [many]; Entrepreneurs; Managers (generic); White collar (clerks, office workers); Blue collar; Agricultural workers; Farmers; Military

Ethnic/racial Divisions [types of]

Age divisions [types of]: Fetuses; Babies; Toddlers; Pre-teens; Teens; Children in general; Adults; Young adults; Middle age; Seniors

People with Disabilities [types of disability] [Coded using linked notation from GA for abilities]

Social Ideology [types of] [May not need to be distinguished from political ideology.]

6. **Politics** [elements of]

Political Institutions [types of]: States (Colonies); Decision-making systems [Governments; Legislatures; Executives; Commissions; Committees; Bureaucracies; Elections, voting; Courts; Demonstrations or protests; Riots; Revolutions or Rebellions; Regulatory agencies]

Rules (documents): Laws; Legal documents; Contracts; Job descriptions; Legal precedents or cases; Forms; Other official documents; Membership forms; Tax forms and regulations; Petitions; Regulations; Licenses; Particular legal decisions (sentences, fines)

Organizations: International; Continental (Regions of countries); National; Regional; State/province; Local; Political parties

Enforcement: Military; Police; Judicial; Prisons

Educational institutions [because they don't fit anywhere else.] [types of]: Universities; Community colleges; High schools; Middle schools; Elementary schools; Schools (general); On-the-job training; Apprenticeship Internship

Political Ideology [types of]: Classical Liberalism; Libertarianism; Pragmatic Liberalism; Conservatism; Socialism; Communism; Fascism; Anarchism; Nationalism

Public Opinion [examples of]: Issues (various) [Cutter numbers might be employed here and in many other places where there are lengthy lists of examples.]

Crime [types of; these will be further subdivided.]: Versus Persons; Versus Property; Civil offences: breach of contract; Torts: breach of duty

7. **Technology and Science** (Scholarship) [Elements of]

[General works or critical appraisals of; Books [many types]; Scholarly articles; Museums; Textual elements in general [Many types]; Libraries [several types]

Theories and Theory Types [Types of, then examples of]: [These are classified in terms of types of agent, action, decision-making, and process, and degree of generalizability.]

Methods [Types of]: Methodology in general (as opposed to methods [Includes measurement, data collection etc.]; Experiments; Surveys; Interviews; Mathematical models; Simulations; Statistical analysis; Ethnographic/observational analysis; Experience/ intuition; Textual analysis; Classification [Several types of each.]

Fields or Disciplines (various) [types of]: History; Geography; Education; Library and Information Sciences; Philosophy; Humanities; Social Sciences; Natural Sciences [several types of each]

Stages of Innovation [types of]: Recognizing the Problem; Setting the Stage; Act of Insight; Critical Revision; Diffusion/transmission [Communication, Adoption]

Innovations [types of] [These will generally be coded with linked notation, and thus we need only list here for certain broad types of innovation: Mechanics; Chemicals; Electrical

8. Health and Population [Elements of]:

Nutrition [Elements of] [Diverse nutritional needs defined physiologically]

Respiration [Needs again defined physiologically]

Disease [Types of]: Viral, bacterial, environmental

Exercise, recreation: [various types]

Sleep

Population

Fertility [Elements of]: Birth; Fecundity; Menarche; Birth intervals etc.

Mortality: Death; Life expectancy; Causes of death [various] [types of] [These can be captured by links to disease, accident, crime/war, and nutrition.]

Migration [types of]: International; Temporary; Nomadic; Permanent; Tourism, Vacation

Age Distribution [types of]

Injuries, Accidents

9. Culture [Elements of]:

Languages [Types of]: Bilingualism; Multilingualism; Dialect [Cutter numbers for particular dialects.]; Animal languages [Linked to particular animals]; Artificial languages (computer languages, created languages); Sign language

[Examples of: ISO 639 provides two-letter codes for 188 world languages; Cutter numbers could be used for others.]

Religions [Elements of]: Formal religions [various]; Providence (the sacred, holy); Revelation (Auspices, dreams); Salvation; Miracles; Doctrine; Conversion; Forms of worship (priesthood, monasticism?); Rituals (baptize, bless, beatify, excommunicate, funeral, burial)

Stories [types of]: Myths; Fairy tales; Legends; Family sagas; Fables and Allegories; Jokes and riddles

Expressions of culture [types of]: Rituals; Dance; Song; Cuisine; Attire; Ornamentation of buildings; Games; Sports [games and sports can be combined with 'children's' or 'individual' or

'competitive' or 'for exercise' Cutter numbers can be used for individual games or sports]; Celebrations or Parties

Values [types of]:

(Goals:) Ambition; Optimism; Attitudes toward wealth; Attitudes toward power; Attitudes toward prestige; Attitudes toward beauty; Attitudes toward honor; Attitudes toward recognition; Attitudes toward love; Attitudes toward friendship; Attitudes toward sex; Attitudes toward incest; Attitudes toward marriage; Attitudes toward physical and psychological wellbeing; Time preference

(Means:) Honesty; Ethics; Righteousness; Attitude toward fate; Work valued intrinsically; Attitudes toward violence; Attitudes toward vengeance; Curiosity; Attitudes toward innovation; Attitudes toward nature; Attitudes toward healing

(Community:) Identity; Family versus community; Openness to outsiders; Trust; Egalitarianism; Attitude to young and old; Responsibility; Authoritarianism; Respect for individuals; Justice; Freedom

(Everyday Norms:) Courtesy; Manners; Proxemics; Tidiness; Cleanliness; Punctuality; Conversational rules; Locomotion rules; Tipping

10. Non-Human Environment: [elements of]:

Location (general coding for) [examples of]: International; Continental (Regions of countries); National; Regional; State/province; Local; Cities; Rural areas; Suburbs; Metropolitan area; Counties or municipalities; Other administrative regions; Boroughs within cities; Neighbourhoods; Highway interchanges; Sites; Geographic features; Bridges and tunnels; Other built structures; Wilderness areas, parks, wildlife refuges; Extinct cities; Ancient kingdoms (other defunct political jurisdictions); Archaeological sites; Disputed territories

Time period (general coding for) [types of]: Temporary or fleeting [Second]; Minute; Hour; Day; Week; Month; Seasonal; Year More than a year; Decade; Century; Millennium; Ongoing; Future; Day/ night cycle; Weekdays; Saturday; Sunday; Holidays; Periodically; Before, preceded by; After, followed by; During/ Now At same time; Again; For the first time; Geological periods

Direction [These can be combined if necessary.]: Central; East; North; South; West; Above or on/below; Inside or intra-, versus outside or inter-,; Near, far, next to, adjacent, beside; Right; Left; Front; Back; Middle; Out of; Toward, at , and away; Up/ down;

Measures: Volume; Distance/length; Number; Area; Mass; Scope/size in general

States of nature [May all get moved into a chemical class]: gas (air, cloud, smoke); liquid (suspension, solution); solid (natural/synthetic, mineral etc.); crystal

Shapes: cube; ball; cylinder; cone; square; circle; hollow; (with) holes; grooved/morticed/guttered; kinked/crooked/bent; granular; folded; wrinkled or ruffled or rumpled or puckered; plumed; pointed versus flat; sharp versus dull; ridge or scallop; sliver/shaving/flaked; twisted; wire or string; gouged or crushed; wave/wavy; curly; embossed; concave versus convex; combed; gnarled; stiff/floppy; layered

Connections: Tributaries; Branches; Fastened; Twisted; Bonded/glued/pasted; Injected; Laminated; Framed; Sewed; Knitted; Lined; Assembled; Doweled; Hinged; Nailed; Screwed; Riveted; Stapled; Tacked; Soldered; Welded; Spliced

Colours Primary, Secondary, Tertiary, Black, white [Cutter numbers for other colours?]

Sounds: loudness; pitch [horn, bell, whistle?]

Smells and Tastes: Sweet/sour; Salty; etc.

Topography: Landforms [Cape; Hills; Island; Mountains; Peninsula; Plateau or plain/ flat land; Steep slopes; Terrace; Isthmus; Valley; Cave]; Natural vegetation [Desert; Forests; Grasslands; Snow-covered]; Water [Canal; Deep water in general vs. Shallow; Fresh; Gulf; Ice; Lake; Dam; Ocean; River; Sea; Strait]; Land and water: [Continental shelf; Delta; Shoreline; Submerged land; Marsh/swamp; Spring; Reef; Seabed; Underground water; Watershed]; Atmosphere [Air; Exosphere; Ionosphere; Stratosphere]; Outer Space [Earth's solar system; Interstellar space]

Climate: Temperature; Precipitation; Air quality; Wind [subclasses of each]

Flora and Fauna; Species (various) [This class is temporary until the natural science schedules are fleshed out.] [Alien and imaginary life forms will be included.]

Materials: Chemical elements; Simple compounds of two elements; More complex natural materials; Artificial materials; Imaginary or hypothesized materials

Natural Resources [It may be best to use linked notation for all subclasses here.]

Natural Disasters [Actually 'natural processes that can be destructive']: Flood; Tornado; Hurricane; Earthquake; Volcano; Blizzard; Fire

Built Environments [Types of]: Offices; Houses; Factories; Fences, walls, etc.; Transport infrastructure [several kinds]; Parks; Farms; Public assembly (hotels, entertainment, store

Artifacts (human) [examples of]

[It may be that this class proves redundant: artifacts can be captured under economic output, art, and perhaps technology.]

Artifacts (non-human) [examples of]: Bird nests, beaver dams, coral reefs; UFOs; etc.