DOCUMENT NUMBER MTR100094

MITRE TECHNICAL REPORT

MITRE

Information Standards and Their Use: Implications for Design Patterns

Dave Mann (G022)
JoAnn Brooks (G022)
16 March 2010

Project No.: 05AAV022-JD:

Table of Contents

1 Introduction

Two Worlds, One Ecosystem

Long before systems engineers began to design computer networks and to integrate databases, there were standards. In fact, the history of standards in fields such as medicine and commerce is measured in centuries. (Bowker & Star, 1999) Later, during the industrial revolution, standards were applied to manufactured goods and we are reminded of the success of such standards when we have a part replaced on our car or when we plug an electrical device into a wall socket. More recently, the worlds of computer systems engineering and standards have become intertwined.

When standards and systems engineering meet, systems engineers are faced with two challenges. First and most obviously, computers, networks, computer programs, and the systems built from them are all manufactured goods. And as such, they are subject to standardization in the same way as other manufactured goods. The challenge for computer and software engineers comes when they are asked to participate in standards efforts. Typically, the act of interacting with other engineers from other organizations and collaborating with them in order to produce standards requires a different set of skills than are needed for effective computer programming and systems engineering. This raises a question that is not typically addressed in their education and training: How are effective standards created?

The second challenge is somewhat more subtle and arguably easier to see only when engaged in enterprise level systems engineering. At the enterprise level, the systems engineer can no longer approach a problem as if there were a single or even small set of primary users for a system. An enterprise is comprised of a large number of smaller sub-organizations. They are related to each other by way of a dynamic and evolving set of relationships, and they must coordinate their actions with each other on an on-going basis. In this context, an enterprise scale system is a part of the communication and coordination work done by the various collaborating sub-groups. That is, the system is part of a socio-technical dynamic that is involved with standardization of the communication and coordination taking place among the cooperating subgroups within the enterprise.

The implication of this for software engineers and data modelers is surprising. The information structures they encode into their systems need to do more than just facilitate information processing by computers. Some of these structures must also encode or create information that can operate as standards for humans engaged in different but related work practices across the enterprise. If it were the case that humans thought and communicated like computers, then traditionally trained software and systems engineers would be well prepared to create such information structures. But the ongoing difficulties in fielding effective enterprise scale information systems indicates that this is not the case. In fact, as technologists who have been shaped by sociology, we assert that this is definitely not the case. Humans do not think like computers. This fact brings us back to a similar but more focused question. How are effective standards for humans created?

While we hope that this paper will be of interest to those who study standards from a socio-technical point of view, our primary intended audience is our fellow technologists and systems engineers who find themselves either dealing with standards efforts directly, or who work to create information systems that must operate at the enterprise scale and facilitate effective standardized communication and collaboration across group boundaries.

We wish to make 3 main points in this paper, which may present something of a bad news/good news theme:

1. The kinds of standards that work for humans are different from the kinds of standards that work well for computers. We expect that this assertion will not be surprising for those from a sociology or socio-technical background. But, for people who are good at creating software systems, this may appear as bad news, as it implies that a different skill set must be brought to bear on the problem (see Section 3).
2. We present a categorization of types of standards that have traditionally worked for humans. While categorizations of this kind are common within the disciplines of computer science and data modeling, to our knowledge, this is the first time such a categorization has been proposed for human oriented standards (see Section 4).
3. Our hope is that the knowledge that 1) human oriented standards are different from machine oriented standards and 2) there is a categorization of known types of human oriented standards will allow system engineers to make more informed design decisions when involved in standards efforts and when creating information systems that must work at the enterprise level (see Section 6).

This paper is organized as follows. In Section 2, we describe an information ecosystem that portrays the communication and collaboration relationships among different but related analytical processes that exist at the enterprise level and that are supported by information systems. For those with a background in socio-technical studies, Section 2 will correctly be seen as a crude caricature. However, our hope is that Section 2 will provide enough of a framework for traditionally trained systems engineers that we will then be able to argue that the social and technical must be considered jointly and, in particular, that human communicative and collaborative needs must be taken into account.

In Section 3, we define the idea of machine oriented standards as those that are created to enable computers to process information. While this section will correctly be seen as a brief summary by those with a computer science or systems engineering background, this section begins to make the argument that machine oriented standards are different from human oriented standards. We argue that the hallmarks of machine oriented standards are: a) their assumption of attribute defined classes, b) stable semantics and c) formalized syntax. Later, we argue that human oriented standards typically do not share these qualities. We also note that machine oriented standards can be categorized according to their structural properties and that this sort of categorization can be used as the basis for creating design patterns. Eventually, we will use this approach to categorize human oriented standards.

In Section 4, we give our categorization of human oriented standards. We identify features of human oriented standards that are not useful for this categorization and those that are. In our discussion of the different types, we also discuss how they fail to possess the three primary characteristics of machine oriented standards.

Having established the primary distinction between human and machine oriented standards in Sections 3 and 4, we consider types of standards associated with the semantic web, such as domain ontologies, and with library and information sciences such as bibliographic indexing systems in Section 5. Despite their similarities to human oriented standards, we note that they possess the three hallmarks of machine oriented standards (attribute defined classes, stable semantics and formal syntax) and that they do so in order to support machine reasoning. For this reason, we argue that they should be considered to be distinct from the human oriented types discussed in Section 4.

In Section 6, we discuss how the categorization of types of human oriented standards might provide a possible basis to form design patterns to facilitate more effective and efficient standards development. In particular, we make an appeal for more ethnographic research to document recurring patterns of communicative and collaboration problems that might be paired with the known types of human oriented standards to create design patterns.

Before proceeding, we note that by attempting to appeal to multiple fields and disciplines, our presentation of this material may use terminology that is either not understood by some readers or that is used in ways that are not expected. For the purposes of this paper, we will define the following terms. More detailed discussions of some of these terms will be given in the following sub-sections.

- SYMBOLIC REPRESENTATION ("REPRESENTATION") - A discrete set of inscriptions (writing, images, graphics, carvings, etc.) that are or become socially recognized as associated with an object or phenomenon.
- ANALYTICAL PROCESS - A process of deliberation followed by a choice or decision; performed by a person, a social group, or a machine, and characterized by input/output relations which accept and produce symbolic representations.
- WRITTEN ARTIFACT - A composite set of symbolic representations (in material and/or electronic form), created by an analytical process and socially recognized as having particular meaning(s).
- INFORMATION PRODUCT - A written artifact that is published and distributed as authorized by a formal organization with the express purpose of being consumed as input by another analytical process or information system.
- STANDARDS BODY - A persistent social group that: a) engages in an institutionalized analytical process, b) produces an information product or products, and c) has recognized authority and legitimacy to establish and regulate standards, including enforcement through application of sanctions against those persons or groups who ignore the standards or use them in illegitimate ways.
- INFORMATION STANDARD ("STANDARD") - An information product that is a) produced by a recognized standards body, and b) is available for broad use among a field, discipline, or industry.
- CLASS - A set of objects or phenomena a) in which all members share a common set of attributes socially recognized as "essential features", b) that has non-ambiguous membership criteria for new objects or phenomena and c) the only possible overlap with other classes is that of superset/subset (e.g., mammals/dogs).
- CATEGORY - A set of objects or phenomena a) in which all members are similar but may not share socially recognized "essential features", b) that may have ambiguous membership criteria for new objects or phenomena and c) that possibly overlap with other categories including nontrivial intersection (e.g., country music and rock-and-roll)
- SET - A class or category that contains members.[1]
- MEMBER - An individual object or phenomenon that belongs to a class or category (i.e., set)
- DEFINITION - A set of symbolic representations or inscriptions within an information product that is recognized by a using community or practice as being associated with a

[1] We recognize that our definitions of "set" and "member" are somewhat circular, and hope that the reader will understand some compromise was necessary here.

specific class, category, or individual object or phenomenon (i.e., member of a class or category).

- LABEL - A concise symbolic representation or inscription within an information product that is recognized by a community of practice as being associated with a definition in the same information product. Because a definition can be associated with a class, category, or member, the label can also be transitively associated with that same class, category, or member.

2 The Information Ecosystem, Systems Engineering and Information Standards

In this section, we describe an information ecosystem that represents the communication and collaboration relationships among different but related analytical processes that exist at the enterprise level. For those with a background in socio-technical studies, this section will likely be seen as a very crude overview of a rich subject area. However, our goal is to provide a framework for traditionally trained systems engineers that will support their developing understanding that the social and technical must be considered jointly.

2.1 Machines Don't Do Judgment, Humans Do

The Basic Analysis Process

The focus of system engineering as it applies to computer systems is often pre-occupied with system integration. A frequent unspoken assumption is that if computers can be made to work together better, then the organizations that use them will be able to work together better. We wish to stand this assumption on its head; beginning, instead, by focusing on the basic human process of analysis as it comes into play in social and professional relationships.

Consider for a moment the process by which a doctor evaluates her patient. Obviously, she will rely on her direct observations of the patient along with her discussion with the patient about the medical issue. The doctor may also review written artifacts including the patient's medical records and reports written by other doctors. But the list of information products that the doctors call upon, either directly or through leveraging past experience, goes beyond those that are directly tied to the individual patient and includes all of the academic and training materials that the doctor has studied, along with all of the credentialing and certification processes the doctor has completed. Some of this information is stored digitally. Some is not. But all of it is brought to bear on the diagnosis problem as the doctor renders her judgment. After the doctor decides upon a diagnosis, she renders it as some form of written artifact. This may be her notes or a formal report. In fact, a whole series of artifacts may result from the doctor's analysis, including additional things such as prescriptions and some of which are published as information products.

This process can be generalized into what we might refer to as the basic human analysis process. Information products are consumed as input synchronically or over time. The human analyst renders an expert judgment. The judgment is captured in some set of written artifacts. These information products are then passed along to other human analytical processes. We do not dispute that there are many other things that are also going on as the doctor proceeds through this process, such as negotiation, identification, distraction, false starts, etc. But for the purpose(s) of this paper, the analytical process is the one currently significant.

2.2 The Real System

Information Products within the Information Ecosystem

This point about information products being passed from one human analytical process to another is critical, because it allows us to recognize that no human analytical process stands alone. Instead, these human analytical processes can be thought of as forming an inter-connected network of analysis processes — an information ecosystem, if you will. The output of one process can serve as input for another. For example, when a doctor produces a report, it is taken as input by the billing department. In like manner, the doctor-produced prescription is taken as input by the pharmacist.

Another point to be made about the information ecosystem (which will be obvious to those with a sociology background but which may be challenging to those with a background in technology) is that the analysis process remains fundamentally dependent upon human judgment. Machines don't render expert judgments; humans do. The roles of the supporting information technologies are to facilitate more efficient storage, retrieval, manipulation, and transport of data. In these ways, the information technology augments the human analytical processes. Rote manipulations of information may be offloaded to computers, but computers do not render expert judgments.

This observation has a corollary that may also be challenging to technologists who are focused on getting computer systems to communicate with each other. Since the core of each analytical process is the human act of rendering a judgment, it then follows that the ultimate consumer of information products passed from one analytical process to the next is a human, not a machine. This is clearly the case when the information product is a non-digital format. For example, a doctor may consult a desk copy of the International Classification of Diseases (a book published by the World Health Organization) or she might hand write a prescription. But this point remains true even when the information products passed from one analytical process to another is in digital form. Data may be processed according to rote rules but the consuming analytical process will in some way render judgments on the information inputs, even if it is nothing more than a human validation that the information was received and processed correctly. A computer generated bank statement would be such a case. While a human analyst at the bank might not analyze each and every bank statement produced, they do analyze the data processing system to ensure that the rendered bank statements meet the goals that they or other humans have established for the bank system. *This evaluative oversight is a human analytical process that cannot be removed, no matter how much automation is applied.*

2.3 Standards within the Information Ecosystem

With the preceding observations in hand, we are now able to place the systems engineering of information systems into a larger context. The system is actually two inter-related networks or eco-subsystems. The first is the human information ecosystem that defines the enterprise or field or discipline. And the second, which is typically the focus of systems engineers, is the network of information technologies used to support the human information ecosystem. One of the primary tenets of socio-technical studies is the assertion that these two systems must be considered as an integrated, comprehensive whole. We fail to understand the human analysis processes if we fail to recognize how the information products and the supporting information technologies augment and otherwise shape the analysis processes. And we fail to understand the information technologies if we fail to recognize the inter-related set of human judgments and communications that it supports.

Communication within the information ecosystem thus often occurs via the sharing of information products (representational artifacts). And among the set of information products that are shared by an enterprise or a field, some are privileged above others, and are considered standards. As we will discuss below, some of these standards are associated with a form of feed-back loop in the information ecosystem.

As an example, in a book titled *Sorting Things Out: Classification and Its Consequences*, Bowker and Star discuss the history of the International Classification of Diseases (ICD), which is produced by the World Health Organization (WHO). The WHO considers, as input, a wide range of information products created by a broad set of analysis processes within the medical community. They consider medical records, academic literature, and a host of other products. As with all human analytical processes, they then render their expert judgments and publish those judgments to create the information product we know as the ICD. But once it is published, the ICD itself is used an input by the very same analytical processes that produce the information products that the ICD takes as input. Hence, a feedback loop is created.

The social theories that figure prominently within the field of socio-technical studies (e.g., ethnomethodology, science and technology studies and structuration theory) offer a variety of ways of understanding this feedback loop. But despite their differences in emphasis, they all agree that standards and their supporting communities of interest co-evolve. And that one possible outcome is for the standards to become more detailed and embedded over time while the supporting communities also become more unified in their practices. As a case in point, Bowker and Star give an excellent recounting of the 300-plus year history of the medical community and its co-evolution with their standardized categorization of diseases and modes of mortality. (Bowker & Star, 1999)

3 Machine Oriented Information Standards

In Section 2, we described an information ecosystem that models the communication and collaboration relationships among different but related analytical processes that exist at the enterprise level. We observed that different analytical processes often communicate with each other using published information products and that some of these information products have the weight of standards.

In this section, we will define the idea of machine oriented standards as those that are created to enable computers to process information. This section begins to make the argument that machine oriented standards are different from human oriented standards. We argue that the hallmarks of machine oriented standards are: a) their assumption of attribute defined classes, b) stable semantics and c) formalized syntax. Later, we will argue that human oriented standards typically do not share these qualities. We also note that machine oriented standards can be categorized according to their structural properties and that this sort of categorization is to be used as the basis for creating design patterns. Later, we will use this approach to categorize human oriented standards.

3.1 An Overview of Machine Oriented Standards

At their core, all computers operate in more or less the same way. A stream of binary data is read in by the central processing unit (CPU) and then separated into two categories. Some of the data is treated as data that is to be stored and manipulated in some way. Other parts of the input stream are recognized as instructions.[2] The CPU has a set of predefined operations that it knows how to perform; and when a binary sequence is read, in that it corresponds to one of these predefined instructions, it performs that operation on data that it has access to.

Obviously, computable binary data streams do not exist on their own — they have to be created by humans. But it is exceedingly difficult for humans to construct such binary sequences directly as a string of 1s and 0s. Instead, humans write their instruction sets in the form of high-level programming languages such as C, Java, and SQL (to name a very few). Concurrent with the programming process, programmers also create formalized data structures that describe the data to be acted on by the program. These include (but are not limited to): arrays, linked lists, stacks, file formats, and databases. These programs and their corresponding data structures cannot be acted upon by the CPU in their native format. Instead, they must be transformed (by computer programs such as compilers) into lower and lower level forms. Eventually, the program and its data are transformed into a pure binary form that can be executed by the CPU.

Having described the basics of computation, we wish to make three observations. First, data structures and programming languages are both subject to standardization at every level. Standards, both in terms of data representation and logical syntax, are mandatory to allow programs to process data and for data to be moved from one computer to another. And these standards are subject to the full range of the social, political, and economic influences identified and discussed by Bowker et al. (Bowker & Star, 1999). Some are closed proprietary standards whose purpose is to guarantee functionality among a controlled set of participants while limiting functionality or control for those outside of the group. Other standards are open, promising interoperability among all who abide by the standard. Some standards are highly formalized in their governance and are managed by international organizations such as the Internet Engineering Task Force (IETF), while others have evolved historically and are not formalized by an identified governing body, yet remain pervasive and broadly accepted none-the-less. For example, the term "stack" refers to a particular form of memory management that is universally understood and affirmed by the computer science curriculum despite the fact that there may be no single identifiable governing body that actively enforces this.

Second, we observe that computing constructs and their corresponding standards exist along a continuum between those types that are directly consumable by the CPU (binary data and CPU instruction sets) and those that allow humans to interact with and control the computer such as formalized data structures and programming languages. *We raise this observation in order to underscore the fact that while these higher level constructs are humanly comprehensible, they remain first and foremost machine oriented in their purpose.* Programming languages with their formalized syntax and data structures with their formalized semantics are formalized precisely because the formalizations allow them to be non-ambiguously translated into binary forms that can be acted upon by the CPU. This is not the way that humans naturally describe objects and actions, nor is it the way that humans tend to communicate with each other, as we describe below.

To make the point very clearly, we stress that even high-level constructs of programming languages are machine oriented. Ultimately, formal data structures and programming languages get compiled and executed according to the rules of binary logic and set theory. It is therefore not

[2] The case of self-modifying code is acknowledged, but not relevant to the current argument.

surprising that classical category theory, rigid semantics, and consistent syntax are the hallmarks of computational standards. The pervasive use of computers and computer networks to solve problems in nearly all aspects of human life demonstrate what can be accomplished using classical category theory, rigid semantics, and consistent syntax. In this way, classically trained computer scientists and programmers are among the most adept at using classical categories, rigid semantics, and formal syntax to model the world. The scope of their success is both far-reaching and undeniable, which leads us to an interesting divergence between the social and cognitive sciences on one hand, and traditionally trained computer scientists on the other. Despite the success of computer science, both cognitive psychologists and social scientists have rejected classical category theory, rigid semantics, and consistent syntax as the way humans think, communicate, and make shared meaning with each other. (Murphy, 2002) (Garfinkel, 2005) (Rawls, 1989) In this light, we recognize that higher level computer languages and data structures remain fundamentally grounded in machine oriented standards. Despite the fact that they are more humanly comprehensible than pure binary forms, the only reason we use them is to express things in a manner that can be ultimately converted to a binary data stream to be executed by a CPU.

Third, we observe that machine oriented standards can be grouped and organized according to what engineers refer to as design patterns. Design patterns are generalized types of solutions that can be applied to similar problems. (Wikipedia, 2010c) Specifics may differ from implementation to implementation, but the basic form of the solution is similar. Pascal, Fortran, and C may be very different in their specifics, but they are all considered to be structured languages. Lisp, Caml, and ML are considered to be functional languages. C++, C#, and Java are object oriented languages. These different language types can be considered to be design patterns. The curriculum of computer science is designed to create proficiency with these different design patterns.

This point is central to the primary thesis of this paper and deserves further expansion. Design patterns can be thought of as being analogous to tools in a tool box. While we know that different screw drivers should be used to drive different types of screws, we also know that all screw drivers are alike in their basic underlying form and, at the same time, they differ in fundamental ways from hammers. In this way, we can recognize a screwdriver and a hammer as distinct design patterns for tools.

Pushing the analogy further, we recognize that nails and screws are the same in that the shared goal of their use is to join pieces of material together. But despite this common goal, screws and nails are very different and are used for different applications. So, the choice between the tools of screwdriver and hammer depends on many factors that ultimately lead to a design decision between the use of screws or nails. If the application demands the use of screws, a hammer will not do. Nor can a screwdriver be used to drive a nail.

Design patterns in computing systems are similar. There may be differences in how one might implement a stack, but when a stack is needed, a linked list will not suffice. Some problem sets are best addressed with object oriented languages while others are better solved using scripting languages.

It is beyond the scope and purpose of this paper to describe the common design patterns used in computer science in detail. For a detailed review of computing design patterns, we refer the reader to the literature of computer science. For the purposes of this paper, it suffices that we establish these three points. First, the programming languages and data structures used for computing are subject to standardization — the results of feedback loops in the information ecosystem of computer science and the field of information technologies. Second, while machine oriented standards vary in the degree to which they are human comprehensible, they all have the hallmark of being ultimately computable. That is, they are built on the concepts of classic category theory, rigid semantics, and formal syntax. Third, standards that have similar structural forms and are well suited to address similar problem sets can be grouped together as design patterns.

4 Human Oriented Standards

In Section 2, we described the information ecosystem in which different but related analytical processes collaborate and communicate with each other via published information products while also being supported by information systems. In Section 3 we defined the idea of machine oriented standards as those that are created to enable computers to process information. We argued that the hallmarks of machine oriented standards are: a) their assumption of attribute defined classes, b) stable semantics, and c) formalized syntax.

In this section, we turn our attention to those standards that shape the information products produced by, and shared among, the human analytical processes within the information ecosystem. It is here that we hope to establish two of our main points: that human oriented standards are different from machine oriented standards; and that human oriented standards can be categorized according to a set of distinguishing features. Developing this second point will be the focus of the sub-sections that follow; so before proceeding with that discussion, we should first discuss how human oriented standards are different from machine oriented standards.

Human oriented standards are rendered in natural languages, not computer programming languages. While natural languages have syntactic rules and norms, their use and application lacks the formality of programming languages. Humans have the ability to comprehend inscriptions (e.g., sentences, definitions) that are not syntactically precise, whereas computers are not. While the fields of natural language processing and natural language understanding remain active areas of computer science research, we note that for computers to manipulate language, it must be converted into data structures that utilize a formal syntax. (Wikipedia, 2010g)

Closely related to this, human language does not only comprise of words that have conventional (invariant) semantic meaning. Some words have the property of indexicality, in which case their meaning can only be understood in the context of its usage. For example, the word “it” has the property of indexicality. (Wikipedia, 2010f) More deeply, language and writing are “resources for action” which people use strategically and interpret according to context in their joint efforts to achieve shared understanding. As Wittgenstein (1953) wrote: “the meaning of a word is its use in the language.” In contrast, a token within a computer system must non-ambiguously refer to a single thing.

Lastly, humans do not conceive of things (categories, objects, phenomena) according to classical category theory. Classical category theory asserts that categories can be defined by a set of central features that are universally shared by all members of the category and that can be used to non-ambiguously make membership decisions. Data elements within computer systems are defined in this manner. However, classification systems have been rejected as a universally best approach for human oriented organizing of information. Experimental results in cognitive psychology confirm that the classical model is not the most common or effective way for humans to understand categories. Gregory Murphy's *The Big Book of Concepts* gives an overview of both the experimentally-based critiques and other conceptual models that are available such the exemplar model and the goal derived model of categories. (Murphy, 2002)

Furthermore, the underlying assumption that things have inherent and discoverable properties has largely been supplanted by understandings of knowledge and language that recognize the necessarily social aspects of shared knowledge. Kuhn and Lakatos offered scientific epistemologies that recognize the role of the community in defining what is considered known according to the overarching paradigm (Kuhn) or research program (Lakatos) and that is upheld and affirmed by peer consensus but that might be overturned or replaced in light of new information. (Wikipedia, 2010e) (Wikipedia, 2010j) Social theorists in Ethnomethodology and Conversation Analysis have argued that categories have meanings that are only understandable in the context in which they are used. (Garfinkel, 2005) (Rawls, 1989) (Sacks, 1995) And the emerging literature on the socio-technical aspects of standards, including works such as Bowker and Star's *Sorting Things Out: Classification and Its Consequences*, demonstrate the social and political influences that shape categorization schemes. (Bowker & Star, 1999) These critiques do not entirely exclude classically defined categories from the realm of human oriented standards. But they do establish that this is not the primary or even preferred way for humans to conceive of concepts[3].

Summarizing, human oriented standards are expressed in human language which lacks the syntactic formality of machine oriented standards. The communication facilitated by human oriented standards relies on contextual or constitutive properties to achieve shared meaning and thus, does not rely on stable semantics in the way that machine oriented standards do. And human oriented standards allow for things to be defined in ways that are not compatible with the classical, attribute based model that is assumed in computer systems.

Having discussed the primary distinction between human and machine oriented standards, we can now turn our attention to developing the categorization of types of human oriented standards. Before describing our categorization, we should make four comments about our approach. First, we will rely heavily on providing examples of the different categories. By drawing on examples that pre-date information technology and the ubiquitous availability of computer systems, we will demonstrate that these types of standards have become well-established patterns because they satisfy human social needs that are distinct from machine computability needs[4]. These types have developed prior to the existence of computers and have developed in ways that are different from machine oriented standards because humans have different intelligibility criteria than Turing machines.

Second, in our description of these different types of human oriented standards, we will attempt to highlight both their structural distinctions and the human analytical process that produce them. In so doing, we hope to increase the understanding among both information system designers and those who study socio-technical systems. For traditionally trained information system designers, we emphasize that *the boundary of the "system" that produces standards must be defined in way that includes the operational human analytical capability that renders the judgments that get encoded in the standard*. For those who approach the problem from a more

[3] A more thorough discussion of classical category theory will be given in sections 4.3.1 and 4.3.2.

[4] A full discussion of what those human social needs are is well beyond the scope of this paper. However, we recognize that a more complete categorization of commonly recurring human needs with respect to doing analysis and communicating between analytical processes will need to be achieved if the types of standards discussed in this section can be used to create fully realized design patterns. (By way of reminder, a design pattern matches an established problem type with an established type of solution.)

traditional sociological starting point, we hope to make the structural distinctions between types of standards more explicit.

Third, the following human oriented categories are not mutually exclusive, nor should they be expected to be. Briefly stated, we are presenting a categorization that admits categories with hazy boundaries and the potential of overlap with other categories (i.e., like musical genres); rather than a rigid classification scheme with no overlapping classes. A more complete explanation of this distinction will be given in the first two sub-sections of Section 4.3.

While we don't appeal to the classical idea of central features that are both necessary and sufficient for defining human oriented standards, different features are more or less useful in drawing distinctions among the types. In Section 4.1, we touch upon features that we believe to be common among human oriented standards and thus not particularly useful with respect to categorizing the types. A more complete discussion of these features is provided in Appendix A. Features that we believe to be of more value in this regard are discussed in Section 4.2. Lastly, in Section 4.3 we present our categorization.

4.1 Common Features of Human Oriented Standards

Human oriented standards have several features that can vary from standard to standard but do not allow us to distinguish between structural types of these human oriented standards. These properties can be thought of as descriptors that might be applied to individual standards efforts. We identify them as being universally applicable across all human oriented standards so as to explicitly exclude them from consideration when distinguishing among types of human oriented standards. We mention three such features:

- **Managerial Formality, Charter, and Enforcement** - Human oriented standards can vary in the degree of managerial oversight, which can range from informal social conventions to highly centralized. They can also vary in terms of the charter of the standards body, typically ranging from proprietary to that of some inter-organizational group, and in terms of enforcement, which can range from none to legally binding.
- **Editorial Posture** - At the core of the work performed by a standards body is the editorial process that establishes the definitions and labels for the things (objects, phenomena, categories, classes) included in the standard. One common editorial distinction is between a descriptive and prescriptive posture.
- **Definitional Detail** - Standards differ in the amount and kind of definitional details used. For example, encyclopedias and dictionaries provide different kinds of detail.

While these features can vary widely across different standards, at this point in time, we do not observe any connection between these differences and the structural type of standards that are used. For this reason, we acknowledge these features but remove them from consideration in our categorization of types of standards. These features are discussed in more detail in Appendix A.

4.2 Distinguishing Properties of Types of Human Oriented Standards

While the features of formality, editorial posture, and definitional detail cannot be used to distinguish between the different types of standards, we assert that the following properties can be used to make these distinctions.

4.2.1 Definitional Focus and Labeling

Some human oriented information standards associate a set of labels (i.e., tokens, names, identifiers) with one or more definitions. Others define things (objects, phenomena, categories, and classes) but lack specific terms to be associated with the definitions. For example, we note that a paragraph in a reference book may make three distinct points, each defining a category. Or, a table might contain five different rows, each describing a different type of thing. But, in both cases, no specific label is associated with the information products. The lack of labels especially becomes clear when attempting to reference those definitions, since lacking any explicit label associated with the information products, we typically default to using implied references based on the presentational structure of the document or information product, such as “the second sentence in the third paragraph of page 48” or “the 5th row of table 8.12”.

In contrast, other standards explicitly associate labels with information products. These labels can take on different forms. Some labels are recognizable as words and are sometimes referred to as *terms* or *proper names*. Examples might include the names of the 50 states in the United States of America or the recognized names of the chemical elements. Other labels combine terms or words in a manner that is meant to convey some descriptive information and these forms of labels are sometimes referred to as *naming schemes*. Examples of naming schemes include the Linnaeus naming system for species, and the conventions for naming chemical compounds. Other labels are composed of alpha-numeric strings that are not commonly recognizable as terms or combinations of terms. These are sometimes referred to as *identifiers* (ids, for short). Identifiers themselves can differ. Some convey some sort of descriptive information while others are strictly nominal. These different approaches to labeling are important but are not sufficient in terms of categorizing human oriented standards. We will use the general term *label* in those cases where the distinctions are not needed.

4.2.2 Structure

Definitions within a standard are related to each other; and the structure of these relationships provides another way that human oriented standards can be distinguished from one another. Adding some complication to this, these relational structures may exist on two different planes: the definitional level and the presentation level. To explain this, we use the example of a dictionary. In the definition of a word, it is common for a dictionary to refer to other related words such as synonyms. This would be an example of a definitional relationship. But we would also say that “apple” and “ant” are both words defined in the “A” section of the dictionary. This statement has nothing to do with the definition of either word. Instead, it is related to the presentational order of the dictionary which is organized alphabetically.

The most complex structure allows for arbitrary relationships between the symbolic representations (either at the definitional or presentational levels) which can be modeled as a *graph* in which the nodes represent the representations and the edges represent the relationships. Definitionally, a dictionary forms a graph. (Princeton University, 2010) However, presentationally, a dictionary forms a rooted tree, a structure we refer to as a *hierarchy*. We note that it is possible for a standard to be hierarchical at the presentation level but not at the definitional level, as the example of the dictionary illustrates. (Obrst L. , Forthcoming)

The explanation of this has to do with the material constraints of presentation. In Library Sciences, this issue is called the shelving problem. A book may be equally justified as being a member of two different subjects simultaneously, yet ultimately, the book must be placed on one shelf. In like manner, the linear aspect of written and printable documents enforces a linear ordering of its subjects, regardless of the relationships among them at the definitional level. Digital media forms such as hypertext can remove this constraint at the presentational level, allowing standards that are conceptually related as graphs to be presented in a non-hierarchical manner. For example, Princeton University's Wordnet provides an English dictionary as a graph structure. (Princeton University, 2010)

Ordered lists are an even simpler structural form in which the only relationship among the items is that given two items, it can be determined that one precedes the other. Addresses along a street or page numbers in a book would be examples of this sort of structure. And lastly, a *set* is the simplest structure that imposes no structural relationship among the items other than inclusion within the set. License plates (within a single state) are an example of a standard that forms a set, since a license plate can be considered valid or invalid (i.e., in the set or not) but among the set of legitimate license plates, there are no defined relationships (apart from possible chronological order of creation).

4.2.3 Accessibility

Standards differ according to the amount of training that is assumed among competent members of the intended usage population. Some are created for general audiences (i.e., broad societies) while others assume specialized knowledge. For example, many campuses (e.g., colleges, hospitals, large companies) utilize some form of a naming scheme to create room numbers which are generally understandable by visitors. In contrast, a repair manual for a jet plane will assume that the reader has a competent understanding of aviation mechanics.

Interestingly, some standards are accessible to different usage populations in different ways. For example, the Dewey Decimal (indexing) system assigns names or IDs to books which encode their subject information. Based upon this information, casual users should be able to locate a particular book in the library, but they would not be expected to determine a book's classification based on the numeric ID; in contrast, a trained librarian is expected to be able to do both.

4.3 A Categorization of Human Oriented Standards

We are now able to present a categorization of different types of human oriented standards. For the first key distinction between classifications and categorizations, we closely follow and summarize the work of Jacob. (Jacob, 2004)

4.3.1 Classifications

Classification Systems, or classifications, form a strictly hierarchical structure at the definitional level and are built on the premise of classical category theory. (Jacob, 2004)Summarizing Jacob, a classification has the following properties:

- Defined by Essential Features - A set of essential features can be identified and stated so that all members of the set share all the essential features.
- Membership Determined by Essential Features - The set of essential features is necessary and sufficient for determining whether or not an object is a member of the set.
- Typicality - All members are equally representative of the set (because all members share all of the essential features).
- Classifier Independence - Individual classifiers will classify the same object in the same way (because membership is entirely defined by the essential features of the object).
- Inheritance - If a class is contained in another class, then its members must possess all of the essential features of the larger class.
- Non-Overlapping Boundaries - Classes that have no containment relationship have no overlap.

- Hierarchical Structure - When classes are related to each other, the relationships are strictly hierarchical.

Classification of this form dates back to the writings of Plato and Aristotle (Wikipedia, 2010b), and the Linnaeus classification system for biological life is often cited as being an example (Wikipedia, 2010h). Introduced in 1735, Linnaeus's system divided nature according to six ranks: kingdom, class, order, genus, species, and variety. The classes within these ranks were determined by observable characteristics. Current biological classification is based on evolutionary history and DNA (Wikipedia, 2010a).

The fact that the biological classification system has been so frequently updated is related to the underlying assumption that things posses inherent properties which, if properly identified, would provide the essential features necessary to construct the corresponding classification system. In this view, an exception to the classification system reveals an error in the classification that demands reconsideration of the corresponding essential features and class definitions. Classification is then taken as a way of constructing real knowledge about the inherent properties of reality. In fact, Jacob notes that

> Until recently, the classical theory of categories exemplified “the ‘right way’ to think about categories, concepts, and classifications” (Gardner, 1987, p. 340). (Jacob, 2004)

That is to say, the use of classification schemes rests on the assumption that reality is inherently classifiable. In this way, there is an implicit assumption that the classification “exists” and the work of the classifier is to “discover” truth by uncovering the correct essential features and class definitions. Closely related to this is the belief that a given subject area possesses a single, canonical classification. However, it is often the case that different analytical points of view will produce different and incompatible hierarchies for the same subject area. As Jacob notes, “The essential observation, however, is that the practice of taxonomy[5] is carried out within the arbitrary framework established by a set of universal principles. ” (Jacob, 2004) That is, different analytical points of view will have different sets of universal principles and the arbitrary choice between these different viewpoints will produce different analytical frameworks and ultimately different classification decisions. Or, as one experienced data modeler has said, “Taxonomies [classifications] are great, if they are your taxonomy. Otherwise, they stink.”

This collision between multiple classification schemes for the same domain is consistent with Garfinkel's ethnomethodological critique of statistics, in which he argued that statistics don’t tell you anything about the world. Instead, they tell you about the accounting practices (or classifications) of the people who produced them. (Garfinkel, 2005) (Rawls A. W., Mann, Garcia, David, & Burton, 2009).

Despite the critiques of ethnomethodology and cognitive psychology as discussed at the beginning of Section 4.3 (Murphy, 2002), classifications remain commonplace. One approach for managing the conflict between the rigid structure of a classic classification and more ambiguous realities is for the classification to become formalized and under the central control of a managing standards body. For example, the Dewey Decimal Classification (DDC) describes a classification system for library books. The first editions of the classification were controlled by

[5] By “practice of taxonomy”, Jacob is referring to the analytical process of classifying an observed phenomenon or object according to a given classification system. In this way, she is using “classification” and “taxonomy” somewhat interchangeably. In contrast, we will make a distinction between these terms below.

its creator, Milvil Dewey. Current editions are now managed by the Online Computer Library Center (OCLC) of Dublin, Ohio. (Wikipedia, 2009b) In like manner, the U.S. Library of Congress maintains its own Library of Congress Classification (LCC). (Wikipedia, 2009f) This is an example of how a single domain can have more than one commonly used classificatory standard and how differences can be managed — each standards body simply manages its own standard separately.

It is also possible for the definition of the classification system to be managed separately from the classification of individual items within that system. For example, while the DDC system itself is directly managed by the OCLC, the assignment of DDC numbers to individual books is done by the Library of Congress (with oversight and review by OCLC). (Wikipedia, 2009b)

Similar to this, it is possible for the constructs of the categorization to be centrally controlled while allowing no central control over the categorization of individual items. In fact, this is typically how biological classifications are managed. For example, the International Code of Zoological Nomenclature (ICZN) is managed by the International Commission on Zoological Nomenclature and provides a classified system of names. But, individual taxonomists work according to the principle of "taxonomic freedom" whereby they can assert new taxonomic classifications and even new taxa (i.e., labels). In this way, the ICZN is more descriptive of common use and less prescriptive than might be popularly understood. (International Commission on Zoological Nomenclature, 1999) (Wikipedia, 2009h) (Encyclopedia Americana, 1991) (Pauly, Hillis, & Cannatella, 2009)

Lastly, another way in which classifications can handle ambiguities is for them to be something less than a perfectly hierarchical classification. A common relaxation of a classification is to allow sub-classes (or items) to be listed as a child (or member) of more than one class at the same time. For example, the DDC provides a mechanism by which the same book can be placed in multiple categories concurrently. One is used for the purpose of shelving while the others are recorded in a subject catalog. (Wikipedia, 2009b)

In terms of definitional focus, we use the term classification to emphasize the relationship among the definitions and to de-emphasize the association between definitions and unique labels. In those cases where each representation is assigned a unique label, we will use the term taxonomy (defined in more detail below), but we note that not all classifications are focused on defining labels in this way. For example, in the United States, the Department of Transportation's requirements for Vehicle Identification Numbers is built upon 4 non-overlapping classes. The first identifies the manufacturer; the second, the type of vehicle; the third is a check digit; and the fourth, the specific vehicle. While these classes are formally defined in Title 49, Section 6, Part 565, Section 25 of the Code of Federal Requirements and while there is an implied labeling at the presentational level (they are defined in subsections (a), (b), (c) and (d) of Sec. 565.25), the definitional focus of Section 565.25 is conceptual, not terminological. For this reason, we would say that the DOT's standard for VINs forms a classification, but not a taxonomy. (U.S. Department of Transportation, 2008)

4.3.2 Categorizations

Again, following the lead of Jacob, we will use the term categorization to refer to standards in which the definitions are related to each other in arbitrary and non-hierarchical ways. We summarize and slightly extend Jacob's discussion of the differences between classifications and categorizations (Jacob, 2004):

- Ambiguous Definitions - Category definitions are formed by changing social influences and may be defined according to a variety (or combination) of cognitive models that defy reduction to a set of essential features
- Flexible Membership - The determination of an object's membership in a category is contingent on the observer's general knowledge and the context of the observation.
- Typicality - Some members are more representative of the category than others.
- Classifier Dependence - Individual categorizers may categorize the same object in different ways.
- Non-Transitive Inheritance - If category A is a subset of B, and B is a subset of C, then A may not be a subset of C. For example, most people will agree that the category car seats is a subset of the category chair while also agreeing that chair is a subset of furniture. But most people will disagree that car seats are a subset of furniture. (Murphy, 2002)
- Overlapping Boundaries - An object may be a member of more than one category at the same time.
- Complex Relational Structure - When categories are related to each other, the relationships are not strictly hierarchical and may form an arbitrary graph structure.

Subject headings used in libraries and music archives are one common example of a categorization. For example, distinguishing between the musical genres of country and western music, or between rock and blues, cannot be adequately reduced to a set of central features. Determination of whether or not a particular work should be considered country or western (or rock or blues) is contingent on the listener's knowledge and the context that the observation is made. Some musical works are considered by many to be more clearly country (or western, or rock, or blues) than others, yet different listeners may categorize the same song differently. Similarly, a musical work might be considered to be both rock and blues at the same time.

It might be argued that the relationship between classically defined classifications and categorizations is that classifications are the goal but categorizations are the reality we are forced to live with until we gain enough knowledge about the subject to create a classical classification. But, the results in both cognitive psychology and social theory of language advanced by EM/CA tell us that some categories cannot be conceived of in terms of essential features and without regard to the constitutive social factors that shape the meaning within particular contexts. That is to say, the inability to create a classification for a domain is not a result of incomplete knowledge, nor is it necessarily the failure of the organizers to find the essential features of the category. In some cases, a categorization scheme is not only the best possible but actually the preferred model for organizing or representing the information, as in the case of the music genre. (Murphy, 2002)

At this point, we are now in a position to clarify an assertion made in the opening of this section — this paper is, itself, a categorization (of types of information standards), not a classification. In particular, we should not expect there to be crisp, cleanly delimited boundaries between classifications and categories. Some systems are clearly classifications, some are clearly categorizations, and some are mixtures of the two. We would assert that most standards that are produced as books or monographs are conceptually structured as categorizations. Since it is common for the concepts defined to overlap and for the relationships between them to form a more complex arbitrary graph, and not a hierarchy, these conceptual relations exist even when the information is presented in a hierarchical manner, such as a document outline.

As was the case with classifications, we use the term categorization in those cases in which the definitional focus of the standard is on the representations themselves, and not a set of labels to be associated with those representations. Standardized procedures or instructions are often examples of categorizations that highlight this lack of definitional focus on labels. The relationships among the steps typically form a non-hierarchical graph, but it is often the case that the individual steps are not assigned any identifying label. In this way, we distinguish between categorizations and dictionaries and controlled vocabularies (which are defined in more detail below). Like categorizations, dictionaries and controlled vocabularies are non-hierarchical in terms of the defined categories, but their definitional focus includes labels and terms, not the categories alone.

4.3.3 Dictionaries, Glossaries and Thesauri

A dictionary is a standard that is definitionally focused on labels (i.e., words or terms) that allows for: a) non-hierarchical relationships among the concepts and b) allows for multiple terms to be associated with the same concept (i.e., synonyms). While both categorizations and dictionaries allow for non-hierarchical relationships among the concepts, we distinguish dictionaries from categorizations when the standard has a clear focus on defining labels (or terms). We further distinguish dictionaries from controlled vocabularies in that controlled vocabularies have the restriction of enforcing a single label for each concept (i.e., no synonyms).

There are two common types of dictionaries. The first is a collection of words and their definitions. The second is a collection of words in one language and their corresponding words in another language. Other examples include glossaries and thesauri. The word *glossary* generally means a listing of defined words that is limited to a field or subject, such as a glossary of boating terms or a glossary for a book. A *thesaurus* is a listing of words that highlights the relationships of a word to other words. In particular, a thesaurus lists words that are similar (synonyms) and words that are of the opposite meaning (antonyms). Definitional dictionaries (or simply dictionaries) and glossaries are the same in that they provide definitions for the words in the list, while inter-language dictionaries and thesauri do not. Alternatively, we might say that inter-language dictionaries and thesauri define terms by way of their relationships to other terms. Definitional dictionaries themselves may differ widely in the kind and amount of definitional detail they provide. For example, etymological dictionaries will emphasize documented usage of words.

Dictionaries are often hierarchical in their presentational structure, but this is subject to change and interpretation. Rendering a dictionary in the printed form of a book creates a shelving problem for the words since an ordering is imposed. While modern dictionaries are arranged alphabetically, this has not always been the case. Early dictionaries were arranged according to subject areas. (Wikipedia, Dictionary, 2010d) When dictionaries are freed from the printed page, the alphabetical or otherwise hierarchical organizational structure may be removed. For example, the WordNet project maintains an English dictionary in the form of a graph. (Princeton University, 2010) Whether organized by subject heading or alphabetically, the hierarchical arrangement of traditional dictionaries creates what Jacob refers to as the “external cognitive scaffolding that provides for the economical storage and retrieval” of the words in the dictionary. (Jacob, Classification and Categorization: A Difference that Makes a Difference, 2004) (Jacob, 2002) That there are multiple hierarchical presentations of words is indicative of the observations that a) the conceptual and presentational structures of a standard may differ, and b) multiple legitimate hierarchical classifications can exist at the same time based on different analytical points of view.

4.3.4 Controlled Vocabularies

A controlled vocabulary is similar to a dictionary with the further restriction of enforcing a single term for every unique definition. In contrast, dictionaries can have words that have multiple meanings (polysemes and homonyms) and different words that are associated with the same concepts (synonyms). (Wikipedia, 2009a)

A typical use for controlled vocabularies is to create a list of terms that have been collected together to facilitate search and retrieval of items (typically documents) within a large corpus. These terms are applied to documents within the corpus (i.e., tagging) and are also used in the categorization or classification structure within a published index for the corpus. (Wikipedia, 2009a) The organized indexes of academic papers that are published by academic organizations, such as the subject headings maintained by the American Mathematical Society (AMS), are examples of this kind of controlled vocabulary. (American Mathematical Society, 2009)

These examples also show that controlled vocabularies might be further distinguished from general purpose dictionaries in terms of their intended audience and their editorial bias. Typically, controlled dictionaries are targeted towards a comparatively unified audience that can be assumed to possess a certain level of expertise both in terms of the subject domain and the purpose of the controlled vocabulary. Closely related to this, there is typically a decidedly prescriptive editorial bias in which the controlled vocabulary is understood to standardize correct usage of terms.

Traditionally, subject headings and formal bibliographic thesauri have been considered to be distinct kinds of controlled vocabularies. *Subject headings* are used by catalogers to describe books or other works of art and tend to use broader terms. In contrast, *(bibliographic) thesauri* are used by indexers to apply search terms to documents and tend to use narrower terms more closely tied to a specific discipline. Thesauri also tend to maintain a richer set of relations between terms, often including “narrower than” and “broader than” type relationships. (Wikipedia, 2009a)

There is a direct line to be drawn from traditional controlled vocabularies and their machine oriented analogs (which will be discussed in more detail in Section 5.1.3). By following a rule of one term for one definition, controlled vocabularies attempt to eliminate semantic ambiguity and, in so doing, they adopt one of the hallmarks of machine oriented standards. We argue that the traditional use of controlled vocabularies to facilitate bibliographic search and retrieval in non-automated corpuses (e.g., academic journals, library card catalogs) is a case in which traditional human practices have anticipated computer enabled search and retrieval. Therefore, it is not surprising that controlled vocabularies are a form of standard that has been used both by humans in an entirely non-automated context and as a form of human/machine interface standard.

4.3.5 Taxonomies

The term taxonomy has many related but different meanings in different fields and is often used interchangeably with the terms classification and hierarchy. (Wikipedia, 2009h) In order to draw clear distinctions between the different structural types of standards we distinguish between the terms as follows. A hierarchy is a structure whose nodes and edges form a rooted tree. A classification defines a set of non-overlapping classes, that are related to each other hierarchically, but does not explicitly assign labels to each class. A taxonomy is a classification that assigns a label or term to each class, that is, while all taxonomies are classifications (the concepts are related hierarchically), not all classifications are taxonomies because they do not have a definitional focus on labels.

For clarity, we reiterate our distinction. We do not define hierarchies as a type of human oriented standard. Rather, hierarchy is a descriptive term that indicates that the members are related in the form of a rooted tree. In contrast, classifications and taxonomies are types of human oriented standards, both of which are hierarchical in terms of their conceptual structure. Classifications and taxonomies differ in terms of their definitional focus. Classifications define only classes (based on the assumption of classically defined categories), whereas taxonomies bind unique labels to each class. In this way, taxonomies can be seen as a proper subset of classifications. All taxonomies are (or are built upon) classifications but not all classifications are taxonomies (when they fail to define labels for the representations of the concepts). The relationship between classifications and taxonomies is analogous to the relationship between categorizations (graphical relationship among representations) and dictionaries and controlled vocabularies (definitional focus on labels and terms).

Taxonomies are traditionally associated with biological names, but have since been applied to many different domains. (Wikipedia, 2009h) A named class within a taxonomy is called a taxon (the plural form is taxa). The rank of a taxon is its level within the hierarchy. For example, most biological taxonomies define the following seven ranks: kingdom, phylum, class, order, family, genus, and species. (Wikepedia, 2009g)

Taxonomies are similar to classified indexes (defined below) in that both are built upon (hierarchical) classifications. However, they differ in terms of which items in the classification are assigned labels. Taxonomies assign names to all taxa at all levels. In contrast, classified indexes only assign labels to individually classified items at the lowest level.

4.3.6 Classified Indexing Systems

A classified index is a system that utilizes an established classification to classify individual items and to assign identifiers to the items that are unique alphanumeric strings which are structured so as to encode the classification judgments. Examples of classified indexing systems include the (US) Library of Congress Classification (LCC) numbers, Dewey Decimal Classification (DDC) numbers, and Vehicle Identification Numbers (VINs). (Wikipedia, 2009b) (Wikipedia, 2009f) (Wikipedia, 2009i)

While a classified index system requires the existence of a classification, we emphasize structure and on-going assignment of identifiers as a separate (but related) standard. For example, the structure of the DDC is managed by the Online Computer Library Center, whereas the assignment of DDC numbers is managed by the Library of Congress. The analytical processes involved are different. Managing the classification structure entails editorial judgments such as balancing descriptive and prescriptive definitions. In contrast, managing the classification process and identifier assignment requires judgments of how to categorize individual items while assuming the existence of the classification structure. Typically, the revision and publication cycle for updates to the classification and the identifiers are on very different time scales. For example, the DDC, LCC, and VIN classifications are updated only occasionally; whereas new DDC numbers, new LCC numbers, and new VINs are published on an on-going basis.

In terms of accessibility, classified indexing systems tend to work on two levels. Information regarding the classification of an object is only comprehensible to those having expert knowledge of both the classification and the encoding system used to produce the id. However, for non-experts, the identifier can still function as a unique label that can be used to confirm identity, even if the descriptive information is not comprehensible. For example, when registering a car, it is common for the owner to present both a title and documented proof of insurance. Both of these documents will typically record the VIN for the car and the clerk will verify that the VINs match. This confirmation of identity is useful but does not require that either the clerk or the owner be able to decode the classificatory information in the VIN.

Similarly, the DDC number is accessible to expert users such as librarians, who can read a DDC number and routinely understand the classification information encoded in it. This is generally not the case for casual users, who use DDC IDs in a strictly nominal (non-descriptive) manner to confirm they have found the correct book or, at most, to find the correct shelf in the library and also which direction to look on that shelf. (Wikipedia, 2009b)

The distinction between VINs and DDC numbers also demonstrates that the encoding may or may not produce identifiers that are recognizably ordered and that correspond to the next thing. Library classifications typically produce such an ordering and this ordering is used to shelve the books, which aids in the retrieval of the book from the shelf since a person can easily determine which direction to look on the shelf for the book. While VINs may also be ordered nominally, this order typically is not used to locate the next car. In so far as some classified indexing systems produced ordered identifiers, we note that there is a non-trivial intersection between this category and that of ordered identifiers, which are defined below.

4.3.7 Ordered Identifier Systems

An ordered identifier system assigns alpha-numeric identifiers to objects in a manner such that the identifiers are recognizably ordered in a way that corresponds to a recognizable ordering of the labeled objects. Common examples include street addresses and serial numbers on manufactured goods. Typically, the ordering of the labeled objects is related to a physical ordering as is the case with addresses or to a temporal feature such as sequence of manufacture.

Room numbers in a building are another example of an ordered identifier system. This example illustrates that, as a type, ordered identifier systems may encode no other information in the identifier beyond order or sequence. At the same time, we note that it is possible for classified indexing systems to be constructed so that the resulting identifiers are ordered and, in this way, we recognize some overlap between the categories. This said, in the general case ordered identifier systems are flat in terms of structural relationships among representations, with no hierarchical and no generalized graphical relationship among the members. The only relationship is that of order.

We note that it is not required for the identifiers to be sequential; only that they be ordered. Nor is it even required that the identifiers be alphabetically or numerically ordered. The requirement is only that the labels have a recognizable order to them. For instance, when presented in the same context, the colors green, yellow, and red are typically understood as forming an ordered set of labels due to their association with traffic lights. (Lampland, et al., 2009) The terms, low, medium, and high form another example.

It may be possible for identifier systems that produce labels that have a recognizable ordering to be interpreted as an ordered identifier system even if that is not the intention. For example, if an organization assigns employees with published employee numbers that can be interpreted as being ordered, the numbers may be interpreted as indicating seniority or rank, even if that is not the case.

Ordered identifier systems may also be referred to as *ordered lists*, *enumerative lists*, or *enumerated lists*. (Wikipedia, 2009e) This usage of the term enumeration is consistent with the convention of enumerated types in programming languages, which is a mechanism that allows programmers to specify an ordered list of values for a variable. (Wikipedia, 2009c) However, we caution that in some contexts, the qualifier enumeration does not imply order. (World Wide Web Consortium (W3C), 2004)

4.3.8 Nominal Identifier Systems

A nominal identifier system assigns unique alpha-numeric identifiers to objects within a single set or category in a manner such that no descriptive information about the individual object is encoded in the identifier. Common examples include: license plates on cars (within a given state), employee or student IDs, and inventory tracking numbers.

While the identifier string, or ID, may contain no characteristic information, it does convey two important forms of information. First, the presence of the identifier asserts that the labeled object is a legitimate and recognized member of the larger set of objects (or category). For example, the presence of a license plate signifies that its vehicle is legitimately registered with the state. An inventory tracking number on a desk chair signifies that it is a part of the managed inventory belonging to the organization. Still, a single object can have multiple identifiers associated with it conveying that it fulfills multiple roles or functions concurrently. For example, the fitness of a car to have a title of ownership that is issued by the state (designated by VINs) and the fitness of a car to be registered for use on the roads (designated by license plates) involve two different sets of analytical decisions by different agencies and organizations. In both cases, the presence of a legitimate identifier signifies that the object has been analyzed by an authorized process and that it has been determined to satisfy the requirements for inclusion. In this way, the assignment and issuance of an identifier by a legitimate process or organization reifies the identity of the object.

While nominal identifiers do not provide any descriptive information about the individual objects, they typically contain some mechanism by which the larger set is identified. For example, it is common for social security numbers to be predicated with the prefix of SSN. Formatting conventions may be enough to indicate the identity of the set. For example, in the US a sequence of 10 digits in the format of (NNN) NNN-NNNN can reasonably be assumed to be a phone number. In still other cases such as license plate numbers, there is nothing in the identifier to indicate the set membership other than presentational context.

The second piece of information conveyed is the uniqueness of the identity. If two identifiers are encountered within information products, an analyst can know that two different objects with discernibly different identities are being discussed. Conversely, if two information products both refer to the same identifier, the analyst is assured they are referring to the same object.

We note that some identifier systems have some amount of descriptive classification information in one part of the identifier and purely nominal characters in another. Social Security Numbers (SSN), International Standard Book Numbers (ISBN), and VINs are examples of such hybrid systems.

We note that nominal identifiers are produced by an organization that manages the analytical capability which assigns and manages the set of approved identifiers. Further, its authority to do so is seen as legitimate and thus the resulting identifiers are recognized as authoritative. Together then, the identifiers, their users, and the administrative authority constitute a socio-technical system. For example, only the Social Security Administration is authorized to produce SSNs. In fact, the production of SSNs by others is treated as a crime.

4.4 Summary

The types of human oriented information standards described in Section 4.3 are summarized in Table 1.

Table 4-1. Human oriented Information Standards

	Example	Definitional Focus	Conceptual Structure	Accessibility
Classifications	Dewey Decimal Classification	Concept	Hierarchy	Expert
Categorizations	Monographs	Concept	Graph	Varies
Dictionaries	Dictionaries	Terms (synonyms)	Graph	Varies
Controlled Vocabularies	Journal Subject Headings	Unique Terms (no synonyms)	Graph	Expert
Taxonomies	Biological Names (Linnaean)	Unique Terms	Hierarchy	Expert
Classified Indexes	Dewey Decimal Numbers	Unique Labels (for classified objects)	Hierarchy	Expert: Detail Casual: Nominal
Ordered Identifiers	Room Numbers	Unique Labels	Ordered List	Casual
Nominal Identifiers	License Plates	Unique Labels	Set	Casual

5 Some Human/Machine Interface Standards

In Section 2, we described an information ecosystem in which human analysis processes (which are supported by information systems) communicate with each other by way of published information products. Both the information system and human information products are shaped by standards, and we have discussed machine oriented and human oriented standards in Sections 3 and 4 respectively. We argued that the hallmarks of machine oriented standards are: a) the assumption of attribute defined classes, b) stable semantics and c) formalized syntax; and we emphasized that human oriented standards do not share these qualities. We also provided a categorization of types of human oriented standards.

We now turn our attention to a set of standard types that sit at the boundary between machine oriented standards and human oriented standards. We focus on two broad areas that may be in the process of merging. The first includes the technologies and forms of the Semantic Web, which include ontologies, formal thesauri, and controlled vocabularies; the second is Library and Information Science (LIS) and the interest in classification and indexing systems for the purpose of automated search and retrieval. Considerable work has already been done in these areas by researchers such as Jacob and Obrst, and we will summarize both of these in the next sections, but we first want to make a cautionary point.

These two fields are both addressing an emerging phenomenon related to the evolution of the information ecosystem over the past ten years: increasingly, the published information products used by human analytical processes to communicate and coordinate with other analytical processes are either being made available in digital format on the World Wide Web, or bibliographic information about them is captured in digital search engines (rather than physical card catalogs). In both cases, the ultimate goal is to use machines to capture, store, search, and retrieve information products which make it easier for humans to find and move information across locations and communities of practice as well as across organizational units (i.e., between different analytical processes).

The major emphasis in both cases is on how information systems can be brought to bear to help communities of humans find and process the information products they need to perform their analysis and judgment work. For both these groups of standards the basic approach and intent is to *codify existing human knowledge* in forms that reflect and support human capacities for creativity — while at the same time being forms that are amenable to storage, transfer, and computation by machines.

Furthermore, because these interface standards types are designed to accommodate many of the same creative capacities that are supported by human oriented standards, many structural similarities and distinctions between the human machine interface standards and human oriented standards are inevitably brought into sharp focus. In fact, the structural similarities are so great that some would argue that these standards types subsume or could be merged with the human oriented standards we discussed in Section 4. Nevertheless, we assert that these human machine interface standards present especially tricky combinations of issues and, therefore, warrant special attention.

As we have explained, human analytical processes and human oriented standards are different than machine based analytical processes and machine oriented standards. In particular, we note that while Semantic Web and LIS-related types of standards have many similarities to those used for human oriented standards, these human machine interface types of standards are designed so that data captured in them can ultimately be compiled into purely digital forms to be manipulated and processed by a CPU. They therefore necessarily have additional levels of formality that render them computable: they are defined using the hallmarks of machine oriented standards: attribute based class definitions, stable semantics, and formalized syntactic representations. For this reason, we ultimately consider them to be a subset of machine oriented standards, at the same time that they retain one leg in the human oriented world. We therefore assert that although there are many similarities between human oriented and human machine interface standards, there are aspects of human oriented standards that cannot be adequately captured by human machine interface standards.

5.1 Semantic Web

The technologies and forms of the Semantic Web include ontologies, formal thesauri, and controlled vocabularies. Here, we will follow the work of Leo Obrst as exemplary. (Obrst L. , Ontological Architectures, Forthcoming) More specifically, the aspirations of the Semantic Web and the emerging fields of study of Knowledge Organization (KO), Knowledge Organization Systems (KOS) and Informatics are to create better formal structures to capture and manage that information. (International Society for Knowledge Organization, 2010) (Wikipedia, 2009d) (Obrst L. , Forthcoming) (Obrst L. , 2003) (Obrst, Semy, & Pulvermacher, 2004)

Obrst will argue (and we agree) that semantic web technologies are used to support machine reasoning (Obrst L. , Forthcoming).

5.1.1 Taxonomies

Obrst defines a taxonomy as an arrangement of terms or representations in a strictly hierarchical manner. He further distinguishes between weak and strong taxonomies. In weak taxonomies the hierarchical relationship between parent and child is purely structural, carrying with it no semantic meaning. Directory structures on file systems are an example. For example, on most computers running Microsoft windows, you will be able to find a directory named “Program Files” and within that directory you will often find another directory named “Windows NT”. Structurally, “Windows NT” is a child of “Program Files” but this doesn’t imply that “Windows NT” is a subset of “Program Files”. In contrast, a strong taxonomy asserts a semantically consistent parent/child relationship in which “Each information entity is distinguished by a property of the entity that makes it unique as a subclass of its parent entity (a synonym for property is attribute or quality).” For example, “Ball Peen” and “Claw” are strict subsets of “Hammer”. The difference between weak and strong taxonomies implies different degrees of machine reasoning that can be supported. In a weak taxonomy (e.g., a directory structure), nothing can be inferred about the relationship between related objects other than the fact that one is the parent and the other the child. In strong taxonomies, the machine can assert semantically meaningful subset inferences. (Obrst L. , Forthcoming)

Both strong and weak taxonomies have counterparts among human oriented standards. We note Orbst's definition of strong taxonomy makes a direct appeal to the idea of essential features that are central to classical category theory as discussed in Section 3.2. Thus, it would be appropriate to consider our definition of human oriented classification schemes and Obrst's definition of strong taxonomies to be equivalent, with a significant caveat. In Section 3.2, we observed that fully realized classification schemes are often an unrealized ideal. Sometimes objects defy classification in a single super-ordinate class and in other cases, the community consensus on organization changes. This means that in the printed form, the classification scheme may provide for exceptions, access to different version of the classification over time, and the adoption of community practices (e.g., taxonomic freedom) to manage ambiguities. As a computational construct, strict taxonomies provide for none of these. Putting this another way, if a human oriented classification scheme is not pure and not stable, a more complex computational model than a strict hierarchy is needed for data modeling.

Weak taxonomies are roughly analogous to human oriented taxonomies. In Section 3.2, we argued that ambiguously defined categorizations (with overlapping categories and contextual membership decisions) are often presented in a hierarchical manner such as in outline form. In so doing, we noted that the true relationships among the categories might form a graph structure and that the imposition of a hierarchical structure was often tied to the physical limitations imposed by the “shelving problem” in its various forms. In a similar manner, Obrst argues that a weak taxonomy is appropriate when, “you just want users to navigate down a hierarchy for your specific purposes”. Examples include user defined directories on computer file systems and menu controls on programs with graphical user interfaces. In this way, we recognize that a weak taxonomy often creates an arbitrarily chosen tree structure that is imposed on a categorization scheme that is more properly conceived of as a graph. As anyone who has reorganized their personal directory (folder) structure on their computer can attest, multiple weak taxonomies might be created for the same data set. In this way, weak taxonomies are largely equivalent to our definition of human oriented taxonomies as discussed in Section 3.2.

Among human oriented standards, (weak) taxonomies are often used when dealing with the imposition of the physical shelving problem such as the need to place a book on one and only one shelf, or the need to organize the sections of a document in a linear outline. Taxonomies as a type for KOS add to this the element of computability and navigation. The parent/child relationships of a taxonomy can be formally captured, allowing the computer to reason regarding the hierarchical structure. However, the computer reasoning is limited to that which can be captured by parent/child relationships. A CPU cannot recognize similarity between two sections of a paper, while it can recognize the hierarchical relationship between them if the section headings are captured in a formal, computable taxonomy (e.g., MS Word's Outline feature).

In our discussion of human oriented taxonomies, we noted that it was often the case that (human oriented) taxonomies are not pure taxonomies. For example, the Linnaeus classification scheme is often cited as an example of a taxonomy, despite the fact that there are species that do not fit neatly in the categorization. In like manner, many weak taxonomies used on computer systems are not true taxonomies. Most file systems allow the creation of cross-references between files (e.g., Microsoft Windows shortcuts or Unix linked files), which admits that some files should be filed in more than one part of the hierarchy at the same time. Likewise, it is common for menu navigation schemes to provide more than one navigation path to the same menu selection. In such cases, we would argue that the presentation scheme encountered by the user is a human oriented taxonomy while the actual computer implementation is using a more complex KOS type, and not a true formal (machine oriented) taxonomy.

5.1.2 Thesauri

In Section 3.2, we discussed human oriented controlled vocabularies that are used for corpus management and noted that traditionally there has been a distinction between subject headings (used to describe books or other works of art and using broader terms) and thesauri (used to apply search terms to documents using narrower terms more closely tied to a specific discipline). Given that the Semantic Web is focused on making web pages more searchable and views web pages as documents, the use of the term thesaurus in the context of facilitating machine supported search and retrieval is consistent with the use of the term “(formal) thesaurus” in traditional print library science. (Wikipedia, 2009a) In this context, the terms thesaurus and controlled vocabulary are synonymous.

The National Information Standards Organization and the American National Standards Institute define a thesaurus as, “a controlled vocabulary arranged in a known order and structured so that equivalence, homographic, hierarchical, and associative relationships among terms are displayed clearly and identified by standardized relationship indicators" (ANSI/NISO, 2005). A machine oriented controlled vocabulary allows for one of five kinds of relationships between objects: a) Synonym or similar to; b) Homonym or spelled the same; c) Broader than or is a superset of; d) Narrower than or is a subset of and; e) Associated or related.

With these caveats noted, we believe there is a functional equivalence between controlled vocabularies in a human oriented context and controlled vocabularies in a KOS setting. In effect, the need to enforce a one to one relationship between terms and definitions in the context of large corpus management pre-dates and anticipates machine assisted search and retrieval.

5.1.3 Ontologies

The term ontology refers to slightly different things in different contexts; but in general, it is associated with the Semantic Web. Leo Obrst notes, “Today's World Wide Web (WWW) is geared toward presenting information to humans. The Semantic Web is an evolution of the WWW that is intended to capture the meaning of data (i.e., data semantics) precisely enough that a software application can interpret them. A key element of the Semantic Web is the use of ontologies to define concepts and their relationships." (Obrst L. , Forthcoming)

Ontologies are data models that allow objects to be related to one another in a non-hierarchical manner. In this manner, ontologies are similar to human oriented categorization systems. They differ however in that they adopt formalities to facilitate machine reasoning. Leo Obrst writes, "as an engineering product is about representing the semantics of the real world in a model that is usable and interpretable by machine." (Obrst L. , Forthcoming) Objects, or entities, are classically defined classes, with non-overlapping boundaries. Syntactically, ontologies are represented in a variety of knowledge representation languages. And semantically, ontologies can encode a range of different amounts of semantics and can be placed along what Obrst refers to as the "ontology spectrum". (Obrst L. , Forthcoming) In the order of increasing semantics, Obrst's spectrum ranges from taxonomies, thesauri, weak ontologies (i.e., conceptual models) and strong ontologies (i.e., logical theories). While Obrst reserves the term "ontology" only for weak and strong ontologies, he argues that they fit along a continuum and uses the term "semantic models" to include ontologies with taxonomies and thesauri. A key distinction that Obrst draws between strong and weak ontologies is that strong ontologies are expressed in knowledge representation languages that are based on formal logic, allowing them to semantically interpret the ontology and to perform "automated reasoning" against it. In contrast, weak ontologies can only read in and process information. Weak ontologies are synonymous with schemas.

Among strong ontologies, Obrst distinguishes between lower (e.g., domain), mid-level, and upper ontologies. A lower, or domain, ontology captures concepts that are associated with a specific community of interest. Mid-level ontologies capture concepts that span domains such as time and location, while upper ontologies contain common sense universal concepts that are shared across all sets of knowledge within a culture. (Obrst L. , Forthcoming)

Obrst also recognizes that ontology designers have many modeling choices to make including: "descriptive vs. revisionary, multiplicative vs. reductionist, universals vs. particulars vs. sets, endurants vs. perdurants, and more". (Obrst L. , Forthcoming) Two of these choices bear mentioning. The descriptive vs. revisionary choice is analogous to the descriptive vs. prescriptive choice facing lexicographers when constructing dictionaries, as we discussed in Section 3.2. The descriptive approach defines concepts as they are used, whereas the revisionary (or prescriptive) approach defines them as they should be used.

The second design choice is between a reductionist posture that attempts to limit the number of concepts to a minimal set of primitives needed and a multiplicative one that allows the addition of multiple related concepts that "can include anything that reality seems to require, and so any distinction that seems useful to make can be made in the ontology." (Obrst L. , Forthcoming) To this we observe that the reductionist approach is similar to the goal of controlled vocabularies (to reduce the number of terms), normalized database models, and mathematical vector spaces (reduce number of spanning vectors). In contrast, the multiplicative approach is similar to the goal of dictionaries, allowing for synonyms and homonyms.

Obrst correctly notes that these two design choices "really have behind them a set of assumptions about how to view the world (e.g., strict realism with no notion of a different possibility)" and notes that reductionist and revisionary approaches are typically found in the same ontologies. To this we would add from an EM/CA point of view, that the necessity of a descriptive and multiplicative approach points to the sociological fact that humans construct sharable categories based on sets of cognitive and social factors that naturally lead to (human-based) standards that do not neatly fit into a reductionist and revisionary framework. That is, socially constructed categories are not created based on strict realism. In this way, we can see ontologies along a spectrum with human oriented categorization schemes ranging from: reductionist/revisionary ontologies to descriptive/multiplicative ontologies to human oriented categorizations.

While ontologies are related to human oriented categorizations, they remain distinct in that their primary purpose is to facilitate machine processing and reasoning. Obrst states, “Today's World Wide Web (WWW) is geared toward presenting information to humans. The Semantic Web is an evolution of the WWW that is intended to capture the meaning of data (i.e., data semantics) precisely enough that a software application can interpret them. A key element of the Semantic Web is the use of ontologies to define concepts and their relationships. With ontologies supplying the context of data, information retrieval and search engines can exploit this contextual information to perform semantic searches.” (Obrst L. , Forthcoming)

5.2 Classification and Library Information Sciences

Library and information science (LIS) shares many of the same goals of the Semantic Web but its scope is broader than web accessible data, and it has special interest in classification and indexing systems for the purpose of automated search and retrieval. We will attempt to follow the work of Elin Jacob as exemplary. (Jacob, 2004) (Jacob, 1991) Jacob argues that LIS formalisms increasingly are called upon to support search and retrieval within the context of information systems (Jacob, 2004)

Jacob quotes Shera (Shera 1960/65) who “has observed that retrieval must be the focus of a theory of library and information science (LIS) and thus ‘the end toward which all our efforts are directed.’” Jacob summarizes Shera's position that the individual and the retrieval system need to be considered at the same time based on three assumptions: “that there are certain cognitive structures that can be identified and described; that it can be demonstrated that these structures are shared across individuals; and that identification of these shared structures will provide the basis for a theory of organization.” (Jacob, 2004)

In this way, we see that search and retrieval systems sit on the boundary between human oriented standards that have shared social recognition and machine oriented standards. They are human oriented in that they provide what Jacob calls the “cognitive scaffolding” used by the searcher. They are machine oriented because they are formalized and implemented in search and retrieval systems.

Toward this goal of search and retrieval, Jacob asserts that a classification system in LIS is “a system of classes, ordered according to a predetermined set of principles and used to organize a set of entities.” (Jacob, 2004) She further states that “Classification as process involves the orderly and systematic assignment of each entity to one and only one class within a system of mutually exclusive and nonoverlapping classes.” (Jacob, 2004)) We wish to highlight several points. First, for Jacob a classification system is based on classical category theory, positing the existence of mutually exclusive classes.

Second, there is a recognition of the human element in the assignment process of entities to these classes. There exists the possibility that different classifiers might classify the same entity differently. In an attempt to counter-act this, the classification scheme and supporting process should be “lawful and systematic: lawful because it is carried out in accordance with an established set of principles that governs the structure of classes and class relationships; and systematic, because it mandates consistent application of these principles within the framework of a prescribed ordering of reality." (Jacob, 2004) There is a parallel here with classification schemes and statistics and we add here that Ethnomethodology and Conversation Analysis (EM/CA) offers a critique of both statistics and the categorization schemes. The EM/CA critique is that in practice, the act of classification tends not to be lawful and systematic. Workers in the field tend to make classifications based on a range of social factors that subvert the stated classification system. (Garfinkel, 2005) (Rawls A. W., Mann, Garcia, David, & Burton, 2009) This critique noted, the goal of systematic and repeatable classification remains the stated goal of LIS classification systems.

Jacob describes four kinds of classification systems: taxonomic classifications, (hierarchical) classification schemes, and bibliographic classifications which she further distinguishes between enumerative classifications and faceted classifications. For Jacob, a taxonomic classification establishes "stability of nomenclature through the aegis of a formalized and universally accepted language". Further, Jacob defines a taxonomic classification as having strict sub-class relationships between parent and children. (Jacob, 2004) In this way, Jacob's definition of a taxonomic classification system is equivalent to Obrst's strong taxonomy as applied to terms and to our definition of human oriented taxonomies.

Jacob's definition of classification schemes is similar to Obrst's definition of strong taxonomies as applied to concepts. Following Shera (1960/65), Jacob asserts that a classification system is a hierarchically arranged set of mutually exclusive classes that can be defined in terms of the classical category theory of essential features. Both are analogous to the (generally unrealizable) ideal of a human oriented classification system as we have discussed it Section 3.2.3.2. In that section, we noted that cognitive psychology and EM/CA both offer critiques that humans generally do not think in these terms. (Murphy, 2002) (Rawls & Mann, Forthcoming) Jacob makes a similar observation when she notes that, "although the classical theory of categories is unable to account for the variability and flexibility of cognitive categorization, it does provide an elegant accounting of the fundamental assumptions on which classification schemes have historically been constructed." (Jacob, 2004)

Lastly, Shera (as quoted by Jacob) defines a bibliographic classification as:

> a list of terms which are specifically and significantly different each from the other, capable of describing the subject content of [resources], inclusive of all knowledge, infinitely hospitable, in an arrangement that is linear, unique, and meaningful, and which when applied to [resources], (Shera 1953/1965)

In terms of logical structure, Jacob's treatment of bibliographic classifications is similar to hierarchical classifications as defined by Jacob or Obrst. But, these bibliographic classifications differ from more general classifications both in terms of application or scope and in the methodology used to create them. In terms of scope, bibliographic classifications have a focused the goal to create labels for bibliographic resources with the presumption of needing to physically store or arrange the resources. This goal has several implications that distinguish bibliographic classifications from more generalized classifications. First, the classification system must be extensible to allow for new classes to be created within the scheme when new resources are encountered that cannot be adequately described in terms of the current classes. It is anticipated at the outset that the classification will need to grow and the classification must permit a process by which new classes can be added in an orderly manner to minimize the impact of these changes.

Second, this physicality or locational aspect of the resources produces the “shelving problem”. Not only must the resource be cataloged, it must be cataloged in a manner that allows the bibliographer to determine where the resource should be shelved. This distinguishes bibliographic classifications from the problems of managing web pages in the Semantic Web. Web pages need globally unique names (provided by URNs and URLs) but these names need not produce a linear ordering of the indexed web pages because web pages do not need to be placed in shelves. Both deductive and faceted classifications produce such orderings, but the orderings are different owing to their different classification methods.

Jacob distinguishes between two different approaches for creating bibliographic classifications: deductive (i.e., top-down) and faceted (i.e., analytico-synthetic). Both are hierarchical in their structure but they differ in how the hierarchy is created. Jacob describes the construction of a top-down classification as involving the “process of division and subdivision of the original universe such that each class, or each level of classes in the structure, is differentiated by a particular characteristic or property (e.g., the property “color” or “shape”).” In contrast, faceted classifications are built bottom-up by identifying similarities among the individual elements and organizing them into mutually exclusive groups. These groups are then organized into larger groups and so on. At the top level, these groups are referred to as facets. (Jacob, 2004)

Both approaches produce labels for individual elements. With top-down approaches, the label is produced as a path from the root to the class containing the element. Faceted classifications produce labels by selecting a single element from each facet and presenting that in an ordered manner. Since both approaches produce hierarchical structures which can be traversed based on a depth first approach, both provided a linear ordering that can be used to shelve items in a linear manner. And both approaches are infinitely hospitable. (Jacob, 2004)

5.3 Summary

The types of standards within Semantic Web/Knowledge Organization and Library and Information Science are very similar. They have similar ways of understanding hierarchical structures like taxonomies. One distinction is that LIS is usually more tied to physical resources that need to be shelved linearly, and as a result, the LIS design types tend to be hierarchical in order to permit a linear ordering of the items. In contrast, Semantic Web design types are more likely to be associated with digital artifacts that do not need to be physically shelved; as a result, they are free to use non-hierarchical structures. In both cases, the design types rely on classic set theory, established semantics, and formalized syntax to enable machine processing and reasoning. While many of these types have direct human oriented analogs, they lack the degrees of freedom to create ambiguity that is useful for humans but problematic for machines. It is this formality, the goal of supporting machine reasoning that the formality supports, that leads us to classify them ultimately as machine oriented standards.

6 Conclusion and Future Directions

We summarize by stating three main points we hope to have established in this paper along with pointing towards future work in the creation of design patterns for human oriented standards.

First, human oriented standards are fundamentally different from machine oriented standards. Human oriented standards are used in an eco-system of interrelated human analytical processes that are co-joined via communicative processes involving the passing of information products that may or may not be in digital form. While information systems can augment human analysis functions and human communications, they cannot replace the essentially social nature of the work performed by humans within the analysis eco-system. In particular, while we recognize the importance of human/machine standards, such as ontologies, schemas, and classified (bibliographic) identifiers (as discussed in Section 5.1), and while we recognize that they have human oriented analogs, we also stress the importance and necessity of the distinct class of human oriented standards, which allow for ambiguity and exceptions.

Second, we have identified some of the major types of human oriented standards and described their primary structural differences, as summarized in Table 4-1. While these terms have commonly been used to describe different types of standards, to our knowledge, no set of structural distinctions among these human oriented standards types has been identified before, and we believe that doing so will assist information systems designers and data modelers in their tasks.

Third, we believe that this categorization of standards may be helpful for those people involved in the authorship of standards; first by making clear the distinction between human and machine orientation; and secondly, by providing a set of design templates for human oriented standards that may be drawn upon by standards designers to suit differing needs.

Our longer range goal is that the types of human oriented standards that we have identified might someday become associated with a set of identifiably recurring communication and analysis problem sets that occur within the analysis eco-system, thereby allowing for the development of design patterns that may be useful for creating standards. Christopher Alexander described the idea of a design pattern by saying, “Each pattern describes a problem that occurs over and over again in our environment, and then describes the core of the solution to that problem, in such a way that you can use this solution a million times over, without ever doing it the same way twice.” (Wikipedia, 2010c)

While our hope is that the types that we have identified will contribute to the development of these design patterns by providing types of solutions, we assume that a categorization of human communication and analysis problems needs to be established first. More generally, we believe that many, if not all, of the different kinds of analytical and communications work utilizing standards within the information ecosystem can be cataloged, and that there are advantages to doing so.[6] Our hope is that the identified patterns of problems could be paired with the defined types of standards that, together, could be used to define useful design patterns.

We acknowledge our posture as technologists influenced by sociology. We understand that the choice of one type of standard over another is not rooted in inherent properties of either the knowledge of a domain or the subjects of its study. Rather, it is rooted in the kinds of the analytical and communicative work being done — how it will be used. We assert that the first question for the designer of standardized information products should not be, “What is the structure and form of this information or subject?” Rather, the first question should be something more along the lines of, “What kind of analytical or communicative work is happening that these products are intended to facilitate?”

[6] For example, we would minimally expect these to include: addressing, confirming identity, searching and browsing.

7 Bibliography

American Mathematical Society. (2009, October). *Mathematics Subject Classi*. Retrieved January 2010, from 2010 Mathematics Subject Classification: http://www.ams.org/mathscinet/msc/msc2010.html

ANSI. (1993). *INISO Z39.19-1993, Guidelines for the Construction, Format, and Management of Monlingual Thesauri.* ANSI.

ANSI/NISO. (2005). *ANSI/NISO Z39.19-2005: Guidelines for the Construction, Format and Management of Monolingual Controlled Vocabularies.* National Information Standards Organization. Bethesda, MD: National Information Standards Organization.

Boden, D., & Zimmerman, D. H. (1991). *Talk & Social Structure.* University of California.

Bowker, G. C., & Star, S. L. (1999). *Sorting Things Out: Classification and Its Consequences.* The MIT Press.

Douglas, M. W., & Clayman, S. E. (2003). Ethnomethodology and Conversation Analysis. In L. T. Reynolds, & N. J. Herman-Kinney (Eds.), *Handbook of Symbolic Interaction* (pp. Chapter 7, pp. 173-202). Lanham, Maryland: Rowman & Littlefield.

Encyclopedia Americana. (1991). Taxonomy. In E. Americana, *Encyclopedia Americana International Edition* (Vol. 26). Grolier Incorporated.

Frohmann, B. (1994). The Social Construction of Knowledge Organization: The Case of Melvil Dewey. *(Advances in Knowledge Organization,Vol. 4*.

Garcia, A. C. (1991). Dispute Resolution Without Disputing: How the Interactional Organization of Mediation Hearings Minimizes Argument. *56:818-835*.

Garcia, A. C., & Jacob, J. B. (1999). The Eyes of the Beholder: Understanding the Turn-Taking System in Quasi-Synchronous Computer-Mediated Communication. *Research on Language and Social Interaction* (32(4):337-367).

Garcia, A. C., & Jacob, J. B. (1998). The Interactional Organization of Computer Mediated Communication in the College Classroom. *Qualitative Sociology* (21(3):299-317).

Gardner, H. (1987). *The Mind's New Science: A History of the Cognitive Revolution.* New York: Basic Books.

Garfinkel, H. (2002). *Ethnomethodology's Program: Working Out Durkheim Aprorism.* (A. W. Rawls, Ed.) Boulder: Paradigm Publishers.

Garfinkel, H. (2005). *Seeing Sociologically: The Routine Grounds of Social Action.* (A. W. Rawls, Ed.) Boulder: Paradigm Publishers.

Garfinkel, H. (1967). *Studies in Ethnomethodology*. Englewood-Cliffs: Prentice-Hall.

Goodwin, C. (1986). Between and Within: Alternative and Sequential Treatments of Continuers and Assessments. *Human Studies* (9(2-3):205-21).

Goodwin, C. Notes on Story Structure and the Organization of Participation. In J. M. Atkinson, & J. Heritage (Eds.), *Structures of Social Action: Studies in Conversation Analysis* (pp. 225-246). Cambridge, England: Cambridge University Press.

Goodwin, C. (1995). The Negotiation of Coherence within Conversation. In M. A. Gernbacher, & T. Givon (Eds.), *Coherence in Spontaneous Text* (pp. 117-137). Amsterdam/Philadelphia: John Benjamins Pub. Co.

Green, R., Bean, C. A., & Hudon, M. (2002). Universality and Basic Level Concepts. *Challenges in Knowledge Representation and Organization for the 21st Century: Integration of Knowledge across Boundaries, Proceedings of the Seventh International ISKO Conference.* Granada, Spain: Ergon Verlag.

Have, P. T. (1999). *Doing Conversation Analysis: A Practical Guide.* London: Sage.

Heath, C., & Luff, P. (2000). *Technology in Action.* Cambridge, England: Cambridge University Press.

Heritage, J. (1987). Ethnomethodology. In A. Giddens, & J. Turner (Eds.), *Social Theory Today* (pp. 224-272). Cambridge, England: Polity Press.

Heritage, J. (1984). *Garfinkel and Ethnomethodology*. Cambridge, England: Polity Press.

Heritage, J., & Atkinson, J. M. (1984). Introduction. In J. M. Atkinson, & J. Heritage (Eds.), *Structures of Social Action* (pp. 1-15). Cambridge, England: Cambridge University Press.

International Commission on Zoological Nomenclature. (1999, May). *Introduction*. (International Commission on Zoological Nomenclature) Retrieved January 2010, from INTERNATIONAL CODE OF ZOOLOGICAL NOMENCLATURE online: http://www.iczn.org/iczn/index.jsp

International Society for Knowledge Organization. (2010, January). *International Society for Knowledge Organization*. (International Society for Knowledge Organization) Retrieved January 2010, from International Society for Knowledge Organization: http://www.isko.org/

Jacob, E. K. (2002). Augmenting Human Capabilities: Classification as Cognitive Scaffolding. *Vol 8*.

Jacob, E. K. (2004). Classification and Categorization: A Difference that Makes a Difference. *LIBRARY TRENDS* (Vol. 52, No. 3, Winter 2004), 515-540.

Jacob, E. K. (1991). Classification and Categorization: Drawing the Line. *Proceeding of the 2nd ASIS SIG/CR Classification Research Workshop.* Washington, D.C.

Jefferson, G. (1974). Error Correction as an Interactional Resource. *Language in Society 2* , 181-199.

Lampland, M., Star, S. L., Bowker, G., Epstein, S., Lengwiler, M., Millerrand, F., et al. (2009). *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life.* (M. Lampland, & S. L. Star, Eds.) Cornell University Press.

Lee, C. P. (2005). Between Chaos and Routine: Boudary Negotiating Artfiacts in Collaboration. *ECSCW 2005: Proceedings of the Ninth European Conference on Computer-Supported Cooperative Work.* Paris, France: Springer.

Luff, P., Hindmarsh, J., & Heath, C. (Eds.). (2000). *Workplace Studies: Recovering Work Practice and Informing System Design.* Cambridge, England: Cambridge University Press.

Maynard, D. W. (2003). *Bad News, Good News: Conversational Order in Everyday Talk and Clinical Settings.* Chicago and London: University of Chicago Press.

Murphy, G. L. (2002). *The Big Book of Concepts.* The MIT Press.

Obrst, L. J., Semy, S. K., & Pulvermacher, M. K. (2004). *Toward the Use of an Upper Ontology for U.S. Government and U.S. Military Domains: An Evaluation.* Bedford, MA: The MITRE Corporation.

Obrst, L. (Forthcoming). Ontological Architectures. *Chapter 2 in Part One: Ontology as Technology in the book: TAO – Theory and Applications of Ontology,Volume 2: The Information-science Stance*.

Obrst, L. (2003). Ontologies for Semantically Interoperable Systems. *Proceedings of the Twelfth ACM International Conference on Information and Knowledge Management* . New York: ACM.

Obrst, L. (2006). *Ontology Spectrum Semantic Models (Power Point).* The MITRE Corporation.

Orlikowski, W. J. (1992). The Duality of Technology: Rethinking the Concept of Technology in Organizations. *Organization Science, Vol. 3* (No. 3).

Pauly, G. B., Hillis, D. M., & Cannatella, D. C. (2009). Taxonomic Freedom and the Role of Official Lists of Species Names. *Herpetologica 65(2):115-128.*

Pomerantz, A. (1984). Agreeing and Disagreeing with Assessments: Some Features of Preferred/Dispreferred Turn Shapes. In J. M. Atkinson, & J. Heritage (Eds.), *Studies in Conversation Analysis* (pp. 57-101). Cambridge, England: Cambridge University Press.
Princeton University. (2010, January). *WordNet*. (Princeton University) Retrieved February 2010, from WordNet: http://wordnet.princeton.edu/
Rawls, A. W. (2001). Durkheim's Treatment of Practice: Concrete Practice vs Representations as the Foundation for Reason. *The Journal of Classical Sociology , 1(1)*.
Rawls, A. W. (1990). Emergent Sociality: A Dialectic of Commitment and Order. *Symbolic Interaction , 13(1)*, 63-82.
Rawls, A. W. (2004). *Epistemology and Practice: Durkheim's Elementary Forms of the Religious Life*. Cambridge: Cambridge University Press.
Rawls, A. W. (2008). Harold Garfinkel and Ethnomethodology. *Organization Studies , 29(5)*, 701-732.
Rawls, A. W. (1989). Language, self, and social order: A reformulation of Goffman and Sacks. *Human Studies 12:147-172*.
Rawls, A. W. (1987). The Interaction Order Sui Generis: Goffman's Contribution to Social Theory. *Sociological Theory , 5 (2)*, 136-149.
Rawls, A. W., & Mann, D. (Forthcoming). *"The Thing is … What is Our 'What'?":An EM/CA Study of a Discussion of "Object" Clarity as a Problem in Information Systems.* The MITRE Corporation.
Rawls, A. W., Mann, D., Garcia, A. C., David, G., & Burton, M. (2009). Simple Enumerations. Ethnomethodology and MITRE Information Assurance Data Standards. *Etnografia e ricerca qualitativa* (Issue 1).
Robinson, E. (2006). *A GUIDE TO NAMES AND NAMING PRACTICES.* UK General Secretariat of Interpol.
Sacks, H. (1992). *Lectures on Conversation.* Blackwell Publishers.
Sacks, H. (1995). *Lectures on Conversation.* Wiley-Blackwell .
Sacks, H. (1984). Notes on Methodology. In J. M. Atkinson, & J. Heritage (Eds.), *Studies in Conversation Analysis* (pp. 21-27). Cambridge, England: Cambridge University Press.
Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-taking in Conversation. *Language , 70*, 696-735.
Schegloff, E. A. (1987). Between Micro and Macro: Contexts and Other Connections. In J. C. Alexander, B. Giesen, R. Munch, & N. J. Smelser (Eds.), *The Micro-Macro Link* (pp. 207-234). Berkely: University of California Press.
Schegloff, E. A. (1988/89). From Interview to Confrontation: Observations of the Bush/Rather Encounter. *Research on Language and Social Interaction , 22*, 215-240.
Schegloff, E. A. (1968). Sequencing in Converational Openings. *American Anthropologist , 70*, 1075-1095.
Schegloff, E. A., & Sacks, H. (1973). Opening up Closings. *Semiotica , 7*, 289-327.
Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The Preference for Self-Correction in the Organization of Repair in Conversation. *Language , 53*, 361-362.
Shegloff, E. A. (1984). On Some Questions and Ambiguities in Conversation. In J. M. Atkinson, & J. Heritage (Eds.), *Structures of Social Action: Studies in Conversation Analysis* (pp. 28-53). Cambridge, England: Cambridge University Press.
Shegloff, E. A. (1987). Recycled Turn Beginnings: A Precise Repair Mechanism in Conversation's Turn-taking Organisation. In G. Button, & J. R. Lee (Eds.), *Social Organisation* (pp. 70-85). Clevedon, England: Multilingual Matters Ltd.

Shera, J. (1960/65). What Lies Ahead In Classification. *Libraries and the Organization of Knowledge* , 129-142.
Star, S. L. (1999). The Ethnography of Infrastructure. *AMERICAN BEHAVIORAL SCIENTIST; Vol. 43 No. 3*.
Star, S. L., & Griesemer, J. R. (1989). Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science* (Vol. 19, 387-420).
Suchman, L. A. (1987). *Plans and Situated Actions: The Problem of Human Machine Communication.* Cambridge, England & New York: Cambridge University Press.
U.S. Department of Transportation. (2008, October 1). *Code of Federal Regulations, Title 49, Volume 6, Chapter V, Part 565.* Retrieved August 2009, from U.S. Goverment Printing Office: http://edocket.access.gpo.gov/cfr_2008/octqtr/49cfr565.25.htm
Whalen, J., & Zimmerman, D. H. (1988). When Words Fail: A Single Case Analysis. *Social Problems , 35(4)*, 335-362.
Wikepedia. (2009g, August). *Taxonomic Rank*. (Wikimedia Foundation, Inc.) Retrieved September 2009, from Wikipedia: http://en.wikipedia.org/wiki/Taxonomic_rank
Wikipedia. (2010a, January). *Biological classification*. (Wikimedia Foundation, Inc.) Retrieved February 2010, from Wikipedia: http://en.wikipedia.org/wiki/Biological_classification
Wikipedia. (2010b, January). *Categorization*. (Wikimedia Foundation, Inc.) Retrieved February 2010, from Wikipedia: http://en.wikipedia.org/wiki/Categorization
Wikipedia. (2009a, August). *Controlled Vocabulary*. (Wikimedia Foundation, Inc.) Retrieved September 2009, from Wikipedia: http://en.wikipedia.org/wiki/Controlled_vocabulary.html
Wikipedia. (2010c, January). *Design Patterns*. (Wikimedia Foundation, Inc.) Retrieved January 2010, from Wikipedia: http://en.wikipedia.org/wiki/Design_pattern
Wikipedia. (2009b, August). *Dewey Decimal Classification*. (Wikimedia Foundation, Inc) Retrieved September 2009, from Wikipedia: http://en.wikipedia.org/wiki/Dewey_Decimal_Classification.html
Wikipedia. (2010d, January). *Dictionary*. (Wikimedia Foundation, Inc.) Retrieved February 2010, from Wikipedia: http://en.wikipedia.org/wiki/Dictionary
Wikipedia. (2009c, December). *Enumerated type*. (Wikimedia Foundation, Inc.) Retrieved December 2009, from Wikipedia: http://en.wikipedia.org/wiki/Enumerated_type
Wikipedia. (2010e, January). *Imre Lakatos*. (Wikimedia Foundation, Inc.) Retrieved January 2010, from Wikipedia: http://en.wikipedia.org/wiki/Imre_Lakatos
Wikipedia. (2010f, February). *Indexicality*. (Wikimedia Foundation, Inc.) Retrieved March 2010, from Wikipedia: http://en.wikipedia.org/wiki/Indexicality
Wikipedia. (2009d, January). *Informatics (academic field)*. (Wikimedia Foundation, Inc.) Retrieved January 2090, from Wikipedia: http://en.wikipedia.org/wiki/Informatics_%28academic_field%29
Wikipedia. (2009e, December). *Library Classification*. (Wikimedia Foundation, Inc.) Retrieved December 2009, from Wikipedia: http://en.wikipedia.org/wiki/Library_classification
Wikipedia. (2009f, August). *Library of Congress Classification*. (Wikimedia Foundation, Inc.) Retrieved September 2009, from Wikipedia: http://en.wikipedia.org/wiki/Library_of_Congress_classification.html
Wikipedia. (2010h, January). *Linnaean taxonomy*. (Wikimedia Foundation, Inc.) Retrieved February 2010, from Wikipedia: http://en.wikipedia.org/wiki/Linnaean_taxonomy
Wikipedia. (2010g, February). *Natural Language Processing*. (Wikimedia Foundation, Inc.) Retrieved March 2010, from Wikipedia: http://en.wikipedia.org/wiki/Natural_language_processing

Wikipedia. (2010i, February). *Oxford English Dictionary*. (Wikimedia Foundation, Inc.) Retrieved February 2/1/2010, 2010, from Wikipedia: http://en.wikipedia.org/wiki/Oed

Wikipedia. (2009h, August). *Taxonomy*. (Wikimedia Foundation, Inc.) Retrieved September 2009, from Wikipedia: http://en.wikipedia.org/wiki/Taxonomy.html

Wikipedia. (2010j, January). *Thomas Samuel Kuhn*. (Wikimedia Foundation, Inc.) Retrieved January 2010, from Wikipedia: http://en.wikipedia.org/wiki/Thomas_Samuel_Kuhn

Wikipedia. (2009i, December). *Vehicle Identification Number*. (Wikimedia Foundation, Inc.) Retrieved December 2009, from Wikipedia: http://en.wikipedia.org/wiki/Vehicle_Identification_Number

Wittgenstein, L. (1953). *Philisophical Investigations.*

World Wide Web Consortium (W3C). (2004, October). *XML Schema Part 2: Datatypes Second Edition: 4.3.5 enumeration*. (World Wide Web Consortium (W3C)) Retrieved December 2009, from W3C: http://www.w3.org/TR/xmlschema-2/#rf-enumeration

Appendix A Common Features of Human Oriented Standards

Human oriented standards have several features that can vary from standard to standard but that exist orthogonal to structural distinctions between types. These properties can be thought of descriptors that might be applied to individual standards efforts. We identify them universally applicable across all human oriented standards, and as significant for some other considerations of human oriented standards. We, therefore, acknowledge them here, yet omit them from our discussion of structural distinctions between types of human oriented standards.

A.1 Managerial Formality, Charter and Enforcement

Human oriented standards can vary in the degree of managerial oversight, which can range from informal social conventions to highly centralized. They can also vary in terms of the charter of the standards body, typically ranging from proprietary to that of some inter-organizational group, and in terms of enforcement, which can range from none to legally binding.

We will use the term *convention* to refer to systems that lack any sort of centralized administrative (analytical and managerial) process. Naming conventions for persons within countries or ethnic groups, which often imply something about the person's lineage, would be one such example. (Robinson, 2006) In some cases, these informal conventions might foreshadow what later becomes recognized as a formalized standard, and in this way, might be seen as something of proto-standard. We will use the term *standard* only in those cases in which there is an organization that is responsible for the production and publication of an information product that has the acceptance and weight of a standard.

Among (formally managed) standards, there can be significant variation in the charter of the managing standards body. We will use the term *proprietary standard* to refer to standards that are managed by an organization for the primary benefit of that organization. For example, organizations that maintain campuses with multiple buildings include building names and room numbers. These standards are managed by some group or groups within the organization and there is little expectation for their use external to the organization. We will use the term *consortium standard* to refer to those standards that are managed by a multilateral set of representatives drawn from a collection of organizations that share a mutual interest. In those cases in which a separate organization is created to have, or is recognized as having, the authority to manage a standard on behalf of a community of interest, we will refer to the resulting standard as a *community standard*. The charter of some standards bodies goes beyond the administration of a single standard or small set of standards to include a large number of standards. In such cases, the standards body may have its own standards on how the standards will be developed and managed. We refer to such standards as *formal standards*. When the standards body is a part of a government, the terms *national, federal, governmental* or *state* might be used as qualifiers. And lastly, when the charter of a standards body is to manage standards that cross country lines and that have specific governmental recognition by participating national governments, we refer to the standards as *international standards*.

Standards can also vary in the degrees to which compliance is audited and enforced. Some standards are not enforced in any formalized manner, relying only social convention. For example, we would recognize the Webster's Dictionary and the Oxford English Dictionary as standards among speakers of the English language, but there is no enforcement mechanism other than social convention.[7] At the other extreme, some standards are enforced with the full weight

[7] Recognizing that educational and publishing institutions do enforce these standards, but not as under penalty of law, we do not include them for the present purposes.

of the law, both in terms of audit and enforcement. For example, false use and representation of Social Security Numbers can result in serious legal consequences. The Linnaeus naming scheme for species is an example of a system that is between these two extremes. The legitimacy of terms used across each of the taxonomic levels is centrally controlled.[8] However, taxonomists are free to apply the principle of “taxonomic freedom” whereby it is legitimate to reclassify a species based on new information, recombining names from the naming scheme (one from each level) in new ways. In this way, it is not surprising to see the same species referred to using different scientific names by contemporary authors reflecting their different theories of the organism’s evolutionary provenance. (International Commission on Zoological Nomenclature, 1999) (Encyclopedia Americana, 1991)

While managerial formality, charter and enforcement vary widely among standards, at this point in time we do not observe any connection between these aspects and the structural type of standards used. For this reason, we exclude this feature from consideration in our categorization of standards types.

A.2 Editorial Posture

At the core of the work performed by a standards body is the editorial process that establishes the definitions and labels for the things (objects, phenomena, categories, classes) included in the standard. One common editorial distinction is between a descriptive and prescriptive posture. A *descriptive* editorial policy aspires to describe things as they are currently used by the supporting community, whereas a *prescriptive* policy aspires to describe them as they should be used, even in those cases where the labels and definitions are not currently used or supported broadly within the supporting community. (Wikipedia, Dictionary, 2010d) Both of these editorial postures are representative of a more fundamental editorial judgment necessarily addressed by all standards, which is the question of inclusion or exclusion from the standard. That is, given the stated scope of a standard, the administrators must make editorial determinations on whether given things (objects, phenomena, categories, classes) should be represented and codified within the standard. In making this decision, the editors are rendering a judgment on whether or not the thing satisfies adequacy criteria to fulfill a role that is recognized within the scope of their effort. (Rawls & Mann, “The Thing is … What is Our ‘What’?”:An EM/CA Study of a Discussion of “Object” Clarity as a Problem in Information Systems, Forthcoming) The descriptive and prescriptive editorial postures are two positions with respect to role adequacy that form a spectrum along which standards efforts vary. However, at this point in time we do not observe any connection between differences in editorial posture and the structural type of standards that are used. For this reason, we also exclude this feature from consideration in our categorization of types of standards.

A.3 Definitional Detail

Standards also differ in the amount and kind of definitional details used. For example, consider the difference between dictionaries and encyclopedias. We might expect to find “dog” defined in both, but the definitions would be very different. In the dictionary, we would expect the definition to be brief and to include classification information (e.g., noun or verb), relationships to other words and etymological information. We would also expect the dictionary to list the definitions for the word as used as both a noun and a verb. In contrast, the encyclopedia would

[8] Albeit regularly updated based on community feedback, as most standards are.

be expected to focus only on canine mammals and to provide different and more detailed information about these animals.

There is also an administrative decision to be made about how to structure the definition. The modern semantic approach is to (attempt to) represent a thing (object, phenomena, category, class) in terms of essential features, its relationship to other defined things (e.g., synonym, antonyms...) and non-ambiguous definitions. The semantic approach is related to the classical view of categories (i.e., words are categories). On the other hand, the etymological approach emphasizes that meaning is only understandable in context and thus, provides examples of proper or current usage. (Wikipedia, Oxford English Dictionary, 2010i)

While the amount and structure of definitions can vary widely across different standards, at this point in time, we do not observe any connection between these differences and the structural type of standards that are used. For this reason, we also exclude this feature from consideration in our categorization of types of standards.