1-1-2011

Differences Between Human Oriented and Machine Oriented Information Standards: Implications for Design of Enterprise-Scale Information Systems

JoAnn Brooks
Syracuse University

David Mann
The MITRE Corporation

DIFFERENCES BETWEEN HUMAN ORIENTED AND MACHINE ORIENTED INFORMATION STANDARDS: IMPLICATIONS FOR DESIGN OF ENTERPRISE-SCALE INFORMATION SYSTEMS

David Mann
THE MITRE CORPORATION

JoAnn Brooks
SYRACUSE UNIVERSITY

Enterprise-scale information systems are deeply entwined with the networks of social practice that use and support them. Yet "interoperability" between information systems and social communities of practice is not always easily achieved, because these disparate types of entities operate according to different logics and respond differently to innovation processes. In this paper we identify differences between the types of information standards used in information systems and those commonly used within social communities of practice, terming the former "machine oriented standards" and the latter "human oriented standards." We then provide a catalog of commonly used human oriented standards. We conclude by suggesting that these distinctions have significant implications for designers and developers of enterprise-scale information standards and information systems.

Acknowledgement: *We thank Deb Bodeau, Richard Games, Martha Lorber, Leo Obrst, Anne Rawls, William Waters and anonymous SIIT reviewers for helpful comments on earlier versions of this paper. Portions of this work were funded through grants from The MITRE Corporation.*

As enterprise systems continue to grow larger, with technical and social aspects becoming more intertwined with each other, information systems design must increasingly take into account both technical and social concerns. Any enterprise-scale technological system must be integrated within a broader-scale organizational and social system that also requires some standardization of communication and coordination across a diverse range of cooperating social groups within the enterprise. Systems engineers working at the enterprise scale can therefore no longer approach a design problem as if there were a single or even small set of primary users for a system, as an enterprise scale information system forms an integral portion of the communication and coordination work done across a wide range of collaborating communities. Rather, designers should assume that data structures that are intelligible for a single set of users may not be universally comprehensible across the enterprise.

And while standards for managing technological interoperability must be formally computable and to some extent remain comprehensible to (knowledgeable) humans, standards for supporting social and organizational coordination adhere to other logics and operate with different parameters. Human analysis generally requires forms of standards (either written or digital) that allow for some degree of ambiguity and judgment and in such cases, rigid data structures that facilitate machine reasoning are ill suited. Yet at the same time, many of the standards that humans rely on in their organizational processes are being rendered in digital form to facilitate processing by computers at some stage. In these latter cases, information artifacts may still be exchanged digitally but are read and analyzed by humans who understand things differently than a computer system does, and make decisions accordingly. The information structures that software engineers and data modelers encode into their systems to support these capabilities therefore need to do more than simply facilitate information processing by computers. Indeed, some of these structures must also encode or create information that can operate as standards for humans engaged in different – but interdependent – work practices across the enterprise. In this paper, we argue that requirements for designing such standards are worthy of serious consideration.

We motivate the design problem by considering the example of the development of The MITRE Corporation's Common Vulnerability and Exposure (CVE) list. CVE assigns nominal identifiers (IDs) to publicly known security flaws in software. The idea for CVE was introduced in 1999, at the 2nd Workshop on Research with Security Vulnerability Databases, hosted at Purdue University (Mann & Christey 1999). While maintainers of vulnerability databases continue to work on how to effectively share data, CVE has enjoyed widespread adoption. As of August, 2011 over 45,000 CVE IDs have been issued, they are used in over 250 products from over 60 countries (MITRE 2011) and are formally recognized by international standards bodies (ITU 2011). We contend that the success of CVE rests on two design decisions that have enabled it to effectively facilitate communication and coordination among humans at the enterprise scale. Both were rooted in the conviction that before the vulnerability management community could agree on how to structure vulnerability information, they must first agree on how to count and identify vulnerabilities. First, vulnerabilities were to be described and recorded using human readable free-text, instead of delineated search attributes as required in a traditional a database. Second, CVE identifiers were to be nominal, containing no hierarchically encoded descriptive information. Both of these decisions succeeded in preventing disputes around the appropriate classification scheme for vulnerabilities. In prior work, we argue that the relative ambiguity of CVE definitions – compared to traditional database entries – allows CVE IDs to operate as realized "boundary objects," facilitating more effective communication and coordination across the boundaries of different but related communities of practice (Rawls, Mann, Garcia, David & Burton 2009).

The very practical goal of this paper therefore is to take a step forward toward assisting designers of information systems and information standards facilitating communication and coordination at the enterprise scale. Our primary point is that the standards that support effective human analysis and communication work differently than those that support machine processing, and that this difference is important. We argue that it is incumbent on designers of enterprise scale information systems and standards to a) recognize situations in which human analysis and communication needs to be supported and b) interweave human oriented design elements with those that are required for effective machine processing where appropriate. This begs two questions. First, what distinguishes human oriented standards from machine oriented standards? Second, can we identify a catalog of the different forms of human oriented standards that designers can leverage for more effective support of human analysis and communication? Answering these two questions is the goal of this paper.

This work reflects the combined efforts of an information systems engineer with a PhD in mathematical graph theory and a growing understanding of ethnomethodology, and a technologist with experience developing user interfaces for large and complex information spaces and a PhD in social science; the paper therefore rests on a somewhat unusual combination of socio-technical foundations. Nevertheless, we believe that the unique combination of our thinking has something worthwhile to offer. Our work originated out of a recurring series of problems that information standards data modelers were having (described above).

Our focus is on human oriented information standards for the management of materially-based objects and phenomena as they are represented in digital form, and we start by distinguishing such human oriented standards from canonical machine oriented standards. In this regard we found the work of information scientist Elinor Jacob (2004) especially helpful, and much of the paper reflects her work on the difference between classification and categorization. We start by describing characteristics of what we term *machine oriented standards* – standards primarily used for interoperability across technical systems. Then we address *human oriented standards* – standards which may be digitally processed, but where the primary analytical processes that concern them are carried out by human beings and social communities. In the remainder of this paper, we describe characteristics of information standards commonly in use within social communities of practice. We highlight how these sets of human oriented standards remain structurally distinct from each other, even while the mechanisms through which they are assigned may be similar. Finally, we consider the implications of these differences for development and evolution of enterprise-scale information systems.

First we begin with some key terms common across machine oriented and human oriented standards:

- **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 (inscriptions).
- **Information Product** - A written or inscribed 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.
- **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 specific class, category, or individual object or phenomenon (i.e., member of a class or category).
- **Structure** - Formal relationships between items referenced by or identified by the standard: how items referenced by a standard are organized within it, and how those differences are represented digitally and conceptually, and how they may legitimately be changed. Following the tradition of graph theory, we identify three main elements of structure: nodes, edges, and labels:
    - **Nodes** – Nodes are logical units of information corresponding to a class or category definition, such as the definition of a word in a dictionary, or an idea defined in a paragraph of text. Nodes also include purely organizational classifications such as a section of a paper, or an entry in a table.
    - **Edges** - Edges define the relationships between the nodes.
    - **Labels** - Labels are names associated with nodes and edges, and are optional. Labels can take on different forms and do not need to be unique. Some labels are recognizable as words and are sometimes referred to as *terms* or *proper names*. 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*. Other labels are composed of alpha-numeric strings that are not commonly recognizable as terms or combinations of terms. These can be referred to as *identifiers* (IDs, for short). Some identifiers convey some sort of descriptive information while others are strictly nominal. We will use the general term *label* in those cases where the distinctions are not needed.[1]

  The simplest structure is a *set* – it requires no characteristic relationship among the items other than commonality of membership within the set. A slightly more structured form is that of *ordered lists*, in which there is a syntax that specifies that given two items, it can be determined that one precedes the other. A more complex structure (with a more complex syntax) allows for arbitrary relationships between set members and can be modeled as a *graph* in which the nodes represent members and the edges represent the

[1] In contrast, other information standards define things (objects, phenomena, categories, and classes) but lack specific terms to be associated with the definitions. The lack of labels especially becomes clear when attempting to reference those definitions, we typically default to using implied references based on the presentational structure of the document or information product, such as "the 5th row of table 8.12" or "the second sentence in the third paragraph of page 48."

relationships. A graph in which any two nodes are connected by one and only one path (i.e. there are no cycles) constitutes a *tree*. A *hierarchy* is a structure whose elements and their relations together form a rooted tree.

The structure of relationships among items referenced within or by an information standard affords a primary characterization of how different types of standards can be distinguished from one another.

- **Category Membership** - The process through which those items or events identified as "similar" are grouped together and paired with a certain meaning or term – how physical items or data instances can be associated with each other, and how it is determined which instances are related to which other instances, and how those relationships are characterized. Differences between human oriented and machine oriented standards exist primarily in this area.

Machine Oriented Information Standards

In this section, we define the idea of machine oriented standards as those information standards that enable computers to process information. We note in particular that machine oriented standards can be categorized according to formal logic and stable semantics.

Overview

For computers to process information, certain base requirements must be met. Information standards comprised of rigid semantics (data representation), logical syntax, and formal rules for category membership, are mandatory to allow programs to process data, and for programs and data to be moved from one computer to another.

We note that each such standard is subject to a full range of social, political, and economic influences (Bowker & Star 1999). Some are closed proprietary standards; others 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.

Yet all machine-oriented standards are structured in terms of syntax and semantics because – 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 with stable definitions (semantics) that is to be stored and manipulated in some way. Other parts of the input stream are recognized as instructions for manipulating the semantic data according to a formal syntax.[2]

These machine oriented standards (i.e. programming languages and data structures) are often further distinguished and categorized according to other structural differences. 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. While the higher level constructs are humanly comprehensible, they remain first and foremost machine oriented in their purpose. Both the formalized syntax of programming languages and formalized semantics of data structures are formalized precisely because this allows them to be non-ambiguously translated into binary forms that can be acted upon by a CPU. Even high-level constructs of programming languages are machine oriented in this way; ultimately, formal data structures and programming languages get compiled and executed according to the rules of binary logic (syntax) and formal set theory (category membership). Thus their structure and category membership conform to the constraints of what is called classical category theory.

Classical Category Theory and Classification Systems

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. Classification of this form dates back to the writings of Plato (2008) and Aristotle (Aristotle & Ackrill 1963; Aristotle & Barnes 1984). 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:

STRUCTURE

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

[2] The case of self-modifying code is acknowledged but remains outside the scope of the current argument.

- *Non-Overlapping Boundaries* – Classes that have no containment relationship have no overlap.

CATEGORY MEMBERSHIP

- *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.
- *Inheritance* – If a class is contained in another class, then its members must possess all of the essential features of the larger class.
- *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).

SUMMARY

To summarize, although higher level computer languages and data structures are more humanly comprehensible than pure binary forms, the primary purpose for their existence remains expressing things in a manner that can be ultimately converted to a binary data stream to be executed by a CPU. Rigid semantics, consistent syntax and classical category theory, are therefore the hallmarks of all computational standards. Structural relationships in machine oriented standards are stated precisely, unambiguously, and non-exceptionally in terms of stable definitions (semantics) and formal syntax. Category membership within machine oriented standards therefore follows the logic of classical category theory (hierarchical set membership). If an instance satisfies the definition, the CPU will accept it as a legitimate member of the class and will process it accordingly.

Human Oriented Standards

We next turn our attention to those standards that shape the information products produced by, and shared among, the human analytical processes within the enterprise. Here we hope to establish two main points: that human oriented standards follow a different implementation (membership categorization) logic than machine oriented standards, and that human oriented standards can also be categorized according to their internal structure and the uses to which they are put.

Human Oriented Standards Differ from Machine Oriented Standards

The standards necessary for human analytical processes are rendered in natural language. And while natural languages have semantics and syntactic rules and norms, their use and application are much less rigidly formal than that of programming languages. Humans have the ability to comprehend written or printed sentences and definitions that are not syntactically precise, whereas computers do not.[3] Furthermore, humans depend upon this capacity as a significant resource in their sensemaking processes. Closely related to this, human language does not only comprise of words that have conventional (invariant) semantic meaning. Many words, such as the pronoun "it," have the property of *indexicality*, i.e. their meaning can only be understood in the context of its usage, where interpretation is a necessarily social process – with participating "others" imagined if not physically present (Weick 1995: 39). More deeply, language and writing are "resources for action" which people use strategically and interpret contextually in their joint efforts to achieve shared understanding. As Wittgenstein (1953) wrote: "the meaning of a word is its use in the language."

Furthermore, humans do not conceive of things (categories, objects, phenomena) according to classical category theory. Experimental results in cognitive psychology confirm that the classical model is not the most common or effective way for humans to understand categories (Murphy 2002). The underlying assumption that things have inherent and discoverable properties–as in classical category theory–has largely been supplanted by understandings of knowledge and language that recognize necessarily social aspects of shared knowledge. Professional communities of practice have a significant role in defining what is "known" according to an overarching paradigm (Kuhn 1962) or research program (Lakatos 1970). It is now widely recognized that what is upheld and affirmed by peer consensus at one point in time might later be overturned or replaced in light of new information. Emerging literature on the socio-technical aspects of standards documents many of the social and political influences that shape categorization schemes (Bowker & Star 1999; Lampland & Star 2009). These

[3] The fields of natural language processing and natural language understanding remain active areas of computer science research; yet for computers to manipulate language, it must ultimately be converted into (and/or from) data structures that utilize a formal syntax.

critiques do not entirely exclude classically defined categories from the realm of human oriented standards; but they do establish that classifications are not the primary or even preferred way for humans to conceive of concepts. Rather, in human to human communication, categories have meanings that are understandable only in the context in which they are used (Garfinkel 2005; Rawls 1989; Sacks 1995).

In other words, communication facilitated by human oriented standards relies heavily on contextual or constitutive properties to achieve shared meaning rather than on that the stable semantics and formal logic associated with machine oriented standards.

Distinctions between Types of Human Oriented Standards

Just as with machine oriented standards, structural properties can be used to distinguish between types of human oriented standards. However, category membership is usually handled differently in human oriented standards than for machine oriented standards. More specifically, category membership processes are socially constituted rather than computationally produced, so that boundaries between the types are hazy and permit overlapping memberships, and category membership processes may even vary within a single type of standard.

A Catalog of Human Oriented Standards

In the sections below, we identify some of the more significant variations on this theme. Our description of the first two structural types (classification schemes and categorizations) again closely follows and summarizes the work of Jacob (2004), and emphasizes the distinction she draws between them.

CLASSIFICATION SCHEMES

Structure

Similar to the use of classification systems in machine oriented standards, the use of classification schemes as human oriented standards rests on the assumptions that reality is inherently classifiable, and that a particular classification "exists," such that 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. Thus, human oriented classification schemes are structured quite similarly to machine oriented classification systems, following the tenets of classical category theory.

Category Membership

It is often the case that different social groups will have different analytical points of view and will produce different and incompatible hierarchies for the same subject area. Thus assignment of labels is managed differently by different communities, and conflicts over classification schemes are not uncommon. For example, when the Linnaeus classification system[4] for biological life was introduced in 1735, it had nature divided according to six ranks: kingdom, class, order, genus, species, and variety, and the classes within these ranks were determined by observable characteristics (Linnaeus 1735). Current biological classification is now based on evolutionary history and DNA studies (Mayr & Bock 2002). The updating of the biological classification scheme reflects underlying assumptions that things possess inherent properties which, if properly identified, would provide the essential features necessary to construct the corresponding classification system. Classification is then taken as a way of constructing real knowledge about the inherent properties of reality, although the views of reality may differ over time and communities.

Although classification schemes are common, how conflicts are managed between different classification schemes and their usage communities, differs across social contexts. One approach for managing the conflict between the rigid structure of a classic classification and more ambiguous realities is to consolidate each classification process under the central control of a single managing standards body and formalize it. For example, although there are two standard classification schemes for library books, each is managed by a separate standards body: the Dewey Decimal Classification (DDC) is managed by the Online Computer Library Center (OCLC) of Dublin, Ohio (OCLC 2003), while the U.S. Library of Congress maintains its own Library of Congress Classification (LCC) (Library of Congress 2011).

Another approach is to manage the definition of the classification scheme separately from the classification of individual items within that system. For example, while the DDC scheme 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) (OCLC 2003). Alternately, the constructs of the categorization may be centrally controlled without sustaining any central control over the categorization of individual items; this is typically how biological

---

[4] We use "classification *system*" here since it is a term in common usage denoting the consistent organization of relationships within the standard. However within our typology, it is more correctly considered a classification *scheme*, since it lacks the rigidly formal category membership mechanism associated with a machine oriented classification system.

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. Yet, individual taxonomists work according to the principle of "taxonomic freedom" whereby they can assert new taxonomic classifications and even new taxa (i.e., labels) (International Commission on Zoological Nomenclature 1999; Wikipedia 2011b; Encyclopedia Americana 1991; Pauly, Hillis, & Cannatella 2009).

Lastly, another way in which ambiguities can be managed within classifications is to allow for less than 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 (OCLC 2003).

CATEGORIZATIONS

We use the term *categorization* to refer to standards in which the definitions are related to each other in arbitrary and non-hierarchical ways, again summarizing and slightly extending Jacob's (2004) work:

Structure

- *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.
- *Complex Relational Structure* – When categories are related to each other, the relationships are not strictly hierarchical and may form an arbitrary graph (network) structure. Furthermore, these relational structures may exist on two different planes: a definitional level and a 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 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. Thus it is possible for a standard to be hierarchical at the presentation level but not at the definitional level (Obrst 2010).
- *Non-Transitive Inheritance* – If category A is a subset of B, and B is a subset of C, A may still 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. Yet most people will disagree that car seats are a subset of furniture (Murphy 2002).

Category Membership

- *Flexible Membership* – The determination of an object's membership within 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.
- *Overlapping Boundaries* – An object may be a member of more than one category at the same time.

Subject headings used in libraries and music archives are one common example of a categorization. For example, distinctions between the musical genres of country and western music, or between rock and blues, cannot be adequately reduced to a set of central features.

Furthermore, human oriented categorization standards also differ according to the amount of training that is assumed among competent members of the usage population. Some are created for general audiences (i.e., broad societies) while others assume specialized knowledge. An example of the former is that 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. An example of the latter is a repair manual for a jet plane, written with the assumption that the reader has a competent understanding of aviation mechanics.

DICTIONARIES

Structure

We distinguish dictionaries from categorizations when the standard has a clear focus on defining labels (or terms). 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. Related types of information standards include glossaries and thesauri.

As noted above, the definitions in a dictionary are commonly related in a *network* form (Princeton University 2010), while the defined terms (labels) are presented in a *hierarchy*. However, this is not necessarily the case, as for example early dictionaries were arranged according to subject areas (Lynch 2005). And when dictionaries are freed from the printed page, the alphabetical or otherwise hierarchical organizational structure may be eliminated altogether, as in the case of the WordNet project which maintains a web-based English dictionary in the form of a linked network (Princeton University 2010).

Category Membership

Category membership reflects usage and decision making by a standards body, typically considered a publisher.

Taxonomies

Structure

The term taxonomy has many related but different meanings in different fields and is often used interchangeably with the terms classification and hierarchy (Wikipedia, 2011a; Wikipedia 2011b). For our current purposes, we distinguish between the terms as follows. Whereas a classification defines a set of non-overlapping classes that are related to each other hierarchically, it does not necessarily have labels assigned to each class. In our typology, a taxonomy is a classification that includes a label or term assigned to each class. Therefore, while all taxonomies are classifications, not all classifications are taxonomies. The relationship between classifications and taxonomies may be considered analogous to the relationship between categorizations (graphical relationship among representations of members) and dictionaries (definitional focus on labels and terms). Taxonomies are traditionally associated with biological names, but have since been applied to many different domains (Wikipedia 2011b). 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 (Wikipedia 2011a).

Indexes

In contrast to taxonomies, which assign names to all taxa at all levels, indexes assign labels only to individually classified items at the lowest level. Most indexes have two tiers of users – common users who are familiar with only the bottom tier (instances), and professional index managers who are also familiar with the mid-levels in the hierarchy and assign labels to the instances. We describe three types of indexes: nominal identifier systems, ordered identifier systems, and classified indexing systems.

Nominal Identifier Systems (Indexes)

Structure

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.

Category Membership

While nominal identifiers do not provide any descriptive information about the individual objects, they typically evidence 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, as a sequence of 10 digits in the format of (NNN) NNN-NNNN can, in the US, be reasonably assumed to be a phone number.

While the identifier string, or ID, may contain no characteristic information, it does convey two important forms of information about how the individual items are being managed. First, since nominal identifiers are produced by an organization that is seen as legitimately managing the analytical capability which assigns and manages the set of approved identifiers, the presence of the identifier asserts that the labeled object is a legitimate and recognized member of the larger set (or category) of authorized objects. For example, an inventory tracking number on an office desk chair signifies that it is a part of the managed inventory belonging to the organization. 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.

Ordered Identifier Systems (Indexes)

Structure

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.[5] As a type, ordered identifier systems may encode no other information in the identifier beyond order or sequence.

However, often alpha-numeric identifiers are assigned to objects within a set 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.

Category Membership

Typically a standards body of some sort manages the assignment of the IDs, to prevent duplication and ambiguity. This might be a facilities department or a zoning commission for example.

Classified Indexing Systems (Indexes)

Structure

A classified index is an information standard that utilizes an established classification to classify individual items and to assign identifiers to those items, using unique alphanumeric strings which are structured so as to encode the classification judgments.

Category Membership

While a classified index system requires the existence of a classification structure, it also requires some system for on-going assignment of identifiers. Examples of classified indexing systems include the (US) Library of Congress Classification (LCC) numbers, Dewey Decimal Classification (DDC) numbers, and Vehicle Identification Numbers (VINs).

In terms of accessibility, classified indexing systems tend to be used by two different levels of communities of practice. Information regarding the classification of an object is comprehensible only to those having expert knowledge of both the classification and the encoding system used to produce the id. Meanwhile, that knowledge is generally opaque to the usage communities. 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. Or as another example, in the US, every car has a vehicle identification number (VIN) which encodes US Department of Transportation identifiers for manufacturer, type of vehicle, and specific vehicle (U.S. Department of Transportation 2005). And while the VIN can be compared and processed as a unique id by motor vehicle registry clerks and insurance agents handling the car’s title and insurance documents, the clerks and agents have little need to know how to decode the classificatory information in the VIN. 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 (above).

Summary

For sensemaking and other knowledge creation endeavors, humans can and need to work with ambiguously-rendered artifacts. While human oriented standards have structure similar to machine oriented standards – i.e. they have nodes, edges and labels – the processes associated with membership categorization are not syntactically pure in human oriented standards. Their logic of category membership is not always distinct, nor rigidly adhered to.

Categorization of human oriented standards is different from machine oriented standards, in the same way that categorization is loose classification – they have overlapping and fuzzy boundaries. However, just because category membership is ambiguous in human oriented standards, does not mean it does not exist. Logics of category membership may be articulated, for example in design documents for human oriented standards, which helps to explain why such a specification may be a “white paper” rather than an ontology.

The types of human oriented information standards described above are summarized in Table 1 (below).

---

[5] The identifiers are not required to be sequential; only ordered. For instance, the colors green, yellow, and red, when presented together in the same context, are typically understood as forming an ordered set of labels due to their association with traffic lights (Lampland & Star 2009). Similarly, the terms, low, medium, and high form another example.

Table 1. *Human Oriented Information Standards*

	Definitional Focus	**Conceptual Structure**	**Example**	**Accessibility**
Classifications	Concept	Hierarchy	Dewey Decimal Classification	Expert
Categorizations	Concept	Network	Monographs	Varies
Dictionaries	Terms (synonyms)	Network (Definitions) + Hierarchy (Presentation)	Dictionaries	Varies
Taxonomies	Unique Terms	Hierarchy	Biological Names (Linnaean)	Expert
Nominal Identifiers	Unique Labels	Set	License Plates	Casual
Ordered Identifiers	Unique Labels	Ordered List	Room Numbers	Casual
Classified Indexes	Unique Labels (for classified objects)	Hierarchy	Dewey Decimal Numbers	Expert: Detail Casual: Nominal

Discussion

We have established three main points in this paper. First, while information technology standards work best for technological interoperability, categories and category membership processes associated with human oriented standards work better for human and social processes. Information systems can augment human communications and social analysis functions, but they cannot entirely replace the interpretive and social nature of the analytical work performed by humans and their social groups within the enterprise. Information standards for human and social analytical processes must therefore necessarily accommodate ambiguity, equivocality and/or indexicality.

And while we recognize the importance of standards intentionally designed to bridge human/machine standards, such as ontologies, data schemas, and classified (bibliographic) identifiers, we note that they remain unambiguous and are consequently not as useful for human sensemaking. We therefore stress the importance and fundamentality of the distinct class of human oriented standards, which support the ambiguity and ad hoc exceptions essential for interpersonal communication and social knowledge creation.

Second, we have identified some of the major types of human oriented standards and characterized their primary structural differences, as summarized in Table 1. To our knowledge, this is the first such catalog of these human oriented standards types, and we claim that the typology is a significant contribution to the literature on information standards.

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. More to the point, just because an artifact is rendered in digital form, does not mean it is a machine oriented standard – and trying to use interface or machine oriented standards to support human communication and collaboration can be problematic. It is therefore important to a) distinguish human oriented standards from machine oriented standards, and b) to identify the best structural design for an information standard depending upon how it is to be used (whether human or machine analytical processes dominate – or some more balanced combination?).

Persons creating new standards should therefore consider how the standards will be used when choosing from among the various forms of human oriented standards. For example, if the community of practice needs to prevent the use of synonyms, a controlled vocabulary (or taxonomy) is needed, not a dictionary. As another example, ordered IDs and nominal IDs allow for different usage: if the users expect an ordered relationship among the labeled items (e.g. exit numbers, room numbers, employees by start date, library indexes) then either a classified index or an ordered index is needed - a nominal identifier will not do, because ordering of labels corresponds to real world geography and is especially useful for "finding things"

We therefore stress that for enterprise-scale system designers, the choice of one standard over another should not simply be based upon assumptions about inherent properties of a particular domain (i.e. classical category theory). Rather, such a choice needs to be grounded in awareness about the kinds of analytical and communicative work being done — i.e. whether such a standard will be used primarily in machine oriented or human oriented analytical processes. The primary question for the designer of standardized information products should be: "What kind of analytical and/or communicative work is happening that these standards are intended to facilitate?"

Limitations and Future Work

A major limitation of this work is that it is illustrated with only a small number of examples. Empirical studies compiling a broader and more robust catalog of human oriented information standards are needed to confirm the accuracy and comprehensiveness of this typology. Secondly, field research to identify patterns of usage against these types, inventorying modalities of human oriented standards usage, would be very helpful toward allowing creators of human oriented standards to design and select more useful human oriented standards.

Ultimately, we would like to see the types of human oriented standards that we have identified here become associated with a set of identifiably-recurring communication and analysis problem sets (including operations such as addressing, confirming identity, searching and browsing). This work might then support the development of design patterns (Alexander 1977; Gamma, Helm, Johnson & Vlissides 1994) for information standards, where a design pattern is comprised of design types plus use case scenarios. These design patterns could then be used to support various types of hybrid socio-technical analysis within enterprises

And perhaps not surprisingly, we also harbor hopes that this work may eventually inform social science theories concerned with socio-technical systems. For example, this work may help clarify some of the asymmetric agency issues intrinsic to actor-network theory, or contribute to knowledge about improving processes of technological innovation.

References

Alexander, C. (1977). *A Pattern Language: Towns, Buildings, Construction*, Oxford University Press.

Aristotle and J. Ackrill (1963). Categories. *Aristotle, Categories and De Interpretatione*. Oxford, At the Clarendon Press.

Aristotle and J. Barnes, Eds. (1984). *The Complete Works of Aristotle: The Revised Oxford Translation*, Princeton University Press.

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

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

Gamma, E., R. Helm, R. Johnson and J. Vlissides. (1994). *Design Patterns: Elements of Reusable Object-Oriented Software.* Addison-Wesley Professional.

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

International Commission on Zoological Nomenclature. (1999, May). Introduction. *International Code Of Zoological Nomenclature*. Retrieved January 2010: http://www.iczn.org/iczn/index.jsp

ITU [International Telecommunication Union]. (2011). ITU-T Work Programme. *International Telecommunication Union*. Retrieved 20 August 2011, from: http://www.itu.int/itu-t/workprog/wp_item.aspx?isn=7359

Jacob, E. K. (2004). Classification and categorization: A difference that makes a difference. *Library Trends* **52**(3), 515-540.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago, University of Chicago Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. *Criticism and the Growth of Knowledge: Proceedings of the International Colloquium in the Philosophy of Science, London, 1965*. I. Lakatos and A. Musgrave. London, Cambridge University Press. **4:** 91-196.

Lampland, M. and S. L. Star, Eds. (2009). *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*, Cornell University Press.

Linnaeus, C. (1735). *Systema Naturae*.

Lynch, J. (2005). How Johnson's Dictionary became the first dictionary. Delivered 25 August 2005 at the Johnson and the English Language Conference, Birmingham, UK. Retrieved 24 April 2011, from http://andromeda.rutgers.edu/~jlynch/Papers/firstdict.html

Mann, David E. and Steven M. Christey (1999). Towards a common enumeration of vulnerabilities. Presented at *The 2nd Workshop on Research with Security Vulnerability Databases*, Purdue University, West Lafayette, Indiana, USA). January 21-22.

Mayr, E. and W. J. Bock (2002). Classifications and other ordering systems. *Journal of Zoological Systematics and Evolutionary Research* **40**: 169–194.

MITRE (2011). CVE Compatibility. *Common Vulnerabilities and Exposures*. The MITRE Corporation. Retrieved 21 August 2011, from http://cve.mitre.org/compatible/

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

Obrst, L. (2010). Ontological architectures. *Theory and Applications of Ontology: Computer Applications.* R. Poli, M. Healy and A. Kameas, Eds. Springer**,** pp. 27-66.

OCLC [Online Computer Library Center] (2003). Introduction to Dewey Decimal Classification.

Orlikowski, W. J. (1992). The Duality of Technology: Rethinking the Concept of Technology in Organizations. *Organization Science,* **3**(3): 398-427.

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

Plato (2008). Statesman. *EBook #1738*, Project Gutenberg. Retrieved 24 April 2011, from http://www.gutenberg.org/files/1738/1738-h/1738-h.htm

Princeton University. (2010, January). WordNet. (Princeton University) Retrieved February 2010, from WordNet: http://wordnet.princeton.edu/

Rawls, A. W. (1989). Language, self, and social order: A reformulation of Goffman and Sacks. *Human Studies 12:147-172.*

Rawls, Anne, Dave Mann, Angela Garcia, Gary David and Matt Burton (2009). Semplici enumerazioni: L’etnometodologia e gli Information Assurance Data Standards del MITRE. (English: “Simple Enumerations: Ethnomethodology and MITRE Information Assurance Data Standards”), *Etnografia e Ricerca Qualitativa* **1**: 77–106.

Sacks, H. (1995). *Lectures on Conversation.* Wiley-Blackwell.

U. S. Department of Transportation, N. H. T. S. A. (2005). "United States Federal VIN Requirements (Title 49, Chapter V, Part 565)." Retrieved 24 April 2011, from http://www.access.gpo.gov/nara/cfr/waisidx_05/49cfr565_05.html

U. S. Library of Congress (2011). "Library of Congress Classification." Retrieved 24 April 2011, from http://www.loc.gov/catdir/cpso/lcc.html

Weick, K. (1995). *Sensemaking in Organizations*. Thousand Oaks, CA, Sage Publications.

Wikipedia (2011a). "Taxonomic rank." Retrieved 24 April 2011, from http://en.wikipedia.org/wiki/Taxonomic_rank

Wikipedia (2011b). "Taxonomy." Retrieved 24 April 2011, from http://en.wikipedia.org/wiki/Taxonomy

Wittgenstein, L. (1953/2001). *Philosophical Investigations.* Blackwell Publishing.