DOCUMENT NUMBER MTR100156

MITRE TECHNICAL REPORT

MITRE

The Relationship between Human and Machine-Oriented Standards and the Impact to Enterprise Systems Engineering

David Mann, JoAnn Brooks & Joe DeRosa

24 February 2011

Table of Contents

1 Introduction

While this paper will provide insights for those who study standards from a socio-technical perspective, our primary intended audiences are: a) systems engineers who become involved in standards efforts and b) data modelers and developers who are designing systems that need to cross organizational boundaries and operate at the enterprise level. These problems are both related to standards.[1]

The problem facing system engineers developing standards is fairly clear. Systems are comprised of many sub-systems and the interfaces that allow them to interoperate with each other are typically codified as standards. The problem here is that the skills needed to produce effective standards are different from those in a traditional engineering education. This paper hopes to provide some guidance to engineers involved in standards efforts in the hopes that they will be better equipped to understand some of the basic trade-offs involved in designing effective standards.

We assert that the problem facing data modelers working at the enterprise level is also a standards related problem, although this may be less obvious. The information captured in these information systems must make sense to people who are members of different but related groups across the enterprise. The analytical work done by these groups is typically done in the context of standards, which includes standardized terms and practices. We are referring here to standards in the broader sense than the purely technical types typically encountered by systems engineers - one that includes things like standardized names for diseases, such as the International Classification of Diseases (ICD) (World Health Organization, 2010), or standardized names for animals (International Commission on Zoological Nomenclature, 1999). Standards of this type predate information systems and the formal standards used to define information systems. For example, the history of the ICD goes back at least to the 1850s (World Health Organization, 2010). We refer to such standards as human-oriented standards because their purpose is to support coordinated human analytical work. And it is here that the problem facing data modelers becomes intertwined with these human-oriented standards. In so far as the information systems designed by the data modeler support the analytical work done by humans across the different groups within the enterprise, they must encode information that reflects, or is based on, the human-oriented standards that help define the analytical work practices of the different groups. That is, there is, or needs to be, a connection between the human-oriented standards recognized by the enterprise and the information systems created by the data modeler. This begs the question: How are human-oriented standards and data models (which are machine-oriented standards) related to each other? This paper hopes to explain some of these relationships.

In our previous work (Mann and Brooks, 2010) we discussed the primary distinctions between human-oriented standards (such as the ICD) and machine-oriented standards (such as computer data models that allow computers to process information). We noted that human analytical work and machine processing of information were interrelated as two parts of a unified information ecosystem in which human analytical processes are supported by information systems and communicate with each other through published information products. In particular, we noted that some of these published information products have the socially recognized status as

[1] We use the term "standards" in a broad sense as any information product that is managed published by a group with the recognized authority to do so. For example, we would consider a software application programming interface (API) to be a standard relative to those groups that wish to utilize it. A more concise definition of standard and standards body are given in the glossary in section 1.1. A more complete discussion can be found in (Mann and Brooks, 2010).

standards and that these standards tend to be associated with a feedback loop that is created when a standards body (itself, an analytical process) takes in published information products in as input and then produces the standard as output, which is then consumed as input by other analytical processes. We also provided a categorization of human-oriented standards based on their differing structural properties in the hopes that those involved with creating standards or creating information systems that support analysis and communication might be able to make more informed design decisions.

In this paper, we explore some of the relationships between machine-oriented and human-oriented standards. We wish to make three primary points:

1. Machine and human-oriented standards exist along a spectrum that is characterized by a fundamental trade-off between the amount of detail encoded in the standard and the amount of diversity among the communities of practice that can access or utilize the standard. We refer to this trade off as the "detail/diversity trade space". In particular, we argue that standards efforts that attempt to gain agreement on too much detail by communities that are too diverse enter what we refer to as "the zone of infeasibility" and become likely to fail. We develop this point in sections 2 and 3.
2. Standards and their supporting communities of practice can co-evolve with each other. In particular, we explore two forms of change. In one, standards can change and become more detailed as the communities' practices become more unified. In the other, standards can solidify and become "locked in" as other standards refer to them. In the first case, we draw on the fields of ethnomethodology and structuration. In the latter, we draw on work in the field of socio-technical studies of standards. We explore these issues in Section 4.
3. We assert that the detail/diversity trade space and co-evolution of standards has direct implications for those involved in the development of information systems, particularly those that operate at the enterprise level and that must cross group boundaries. We discuss several of these implications in Section 5.

1.1 Some Terms Used in This Paper

We recognize that many of the terms used in this paper have subtle but important differences in their implications across different fields. While we do not aspire to create a glossary that will satisfy all fields, we present several of the key terms as they will be used in this paper.

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 (i.e., 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 in which a) all members share a common set of attributes, socially recognized as "essential features," b) have 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) the possible overlap with other categories includes nontrivial intersection (e.g. country music and rock 'n' roll)

SET - A class or category that contains members.[2]

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 of practice as being associated with a 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 within 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.

DEFINITIONAL FOCUS (OR FOCUS) - The referent of the definitions and labels (i.e., inscriptions) within an information product including: *distinct classes or categories*, in which different classes or categories are clearly defined despite the fact that the definitional inscriptions are not clearly distinct (e.g., several categories defined in a single paragraph of text); *distinct definitions*, in which the definitional inscriptions are clearly delineated from each other (e.g., bulleted lists, or table entries) and; *distinct labels*, in which unique labels are clearly associated with distinct definitions.

DEFINITIONAL STRUCTURE (OR STRUCTURE) - The structural relationship among the definitions within an information product (or a portion of an information product) including: *graphical*, in which the relationships among the definitions form an arbitrary graph; *hierarchical*, in which the relationships among the definitions form a rooted tree or hierarchy; *ordered set*, in which the defined classes, categories, or members are ordered; and *(unordered) set*, in which the defined classes, categories, or members have no explicit or implicit ordering relationship among them.

MACHINE ORIENTED STANDARDS - Standards that are created to enable computers to process information. The hallmarks of machine-oriented standards are: a) their assumption of attribute defined classes, b) stable semantics and c) formalized syntax. Examples include (but are not limited to): programming languages, data file formats, and shared executable files.

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

HUMAN ORIENTED STANDARDS - Standards that are created to facilitate communication and collaboration among a group of related human analytical processes. They are rendered in natural languages with the assumption of human readership. Examples include (but are not limited to): classifications, dictionaries, ordered identifiers, and nominal identifiers.[3]

INTERFACE STANDARDS - Standards that are created to capture human-oriented information products (a.k.a. knowledge) in a digital format to facilitate machine processing and manipulation of the knowledge, including search and retrieval. Interface standards share the hallmarks of machine-oriented standards: namely: a) the assumption of attribute defined classes, b) stable semantics and c) formalized syntax. Examples include (but are not limited to): schemas, ontologies and bibliographic indexes.

CLASSIFICATION - A type of human-oriented standard in which a) the non-overlapping classes are defined by a set of essential features, b) definitions are related to each other in a hierarchical manner but, c) the classes are not associated with defined labels (or names).

CATEGORIZATION - A type of human-oriented standard in which a) the categories are ambiguously defined and may overlap, b) definitions can be related to each other in an arbitrary (non-hierarchical) manner and, c) the categories are not associated with defined labels (or names).

DICTIONARY - A type of human-oriented standard in which a) the categories are ambiguously defined and may overlap, b) definitions can be related to each other in an arbitrary (non-hierarchical) manner, c) the categories are associated with defined labels (or names) and, d) multiple labels can be associated with the same concept (i.e., synonyms).

TAXONOMY - A type of human-oriented standard in which a) the non-overlapping classes are defined by a set of essential features, b) definitions are related to each other in a hierarchical manner and, c) the classes at each level of the hierarchy are associated with defined labels (or names).

jCONTROLLED VOCABULARY - A type of human-oriented standard in which a) definitions can be related to each other in an arbitrary (non-hierarchical) manner, b) the categories are associated with defined labels (or names) and, c) a restriction is enforced that guarantees a single unique label is associated with each definition (i.e., synonyms are not allowed). Similar to the human machine interface type of (formal) thesaurus (a.k.a., formal controlled vocabulary).

CLASSIFIED IDENTIFIERS - A type of human-oriented standard 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 United States (U.S.) Library of Congress Classification (LCC) numbers, Dewey Decimal Classification (DCC) numbers and Vehicle Identification Numbers (VINs). Similar to the human machine interface type of bibliographic indexing systems.

ORDERED IDENTIFIERS - An ordered identifier system is a type of human-oriented standard that 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.

[3] A more comprehensive review of human-oriented standards and their distinction from machine-oriented standards is given in our prior work. (Mann, et al., 2010) (Mann, et al., 2010)

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.

NOMINAL IDENTIFIERS - A nominal identifier system is a type of human-oriented standard that 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, employee or student IDs, and inventory tracking numbers.

2 Standards Viewed in Three Spectrums

In our previous work (Mann, et al., 2010) we distinguished between three primary categories of standards: standards that support human analysis and communication; standards that support machine reasoning and those that sit on the human/machine interface. We observe that the three categories and the standard types within them can be viewed as existing along three different spectrums.

2.1 Intended Audience

The first spectrum is the intended audience which ranges from human-oriented to machine-oriented. A computer's central processing unit (CPU) is the primary component responsible for information processing, and it can only process highly structured binary files. While these files are ultimately created by humans (typically using a variety of technologies), they exist at one polar extreme of the human/machine spectrum. Generally speaking, humans do not interact directly with pure binary files. Instead, humans interact with data in more humanly accessible forms, which are translated into, and out of, pure digital formats. For example, it is common for a program to be written in what is referred to as a third generation language such as C, Java, or Perl, which is more humanly, comprehensible and which allows the programmer to specify both the form of data structures and the processing logic of the program. Once the program is complete, it can be translated or compiled into more machine-oriented, but still humanly comprehensible types such as assembly language. After that, the assembly code can then be converted or linked to create the machine executable binary type of the program. Similar distinctions exist among data files that are processed by programs on a computer. Some data file standards encode the data in structured binary formats that are not meant to be read by humans, while others, such as mark up languages like XML and HTML, are meant to be readable by expert human readers.

We summarize the spectrum of machine-oriented standards as they range from being more machine-oriented to more human-oriented in Figure 1.

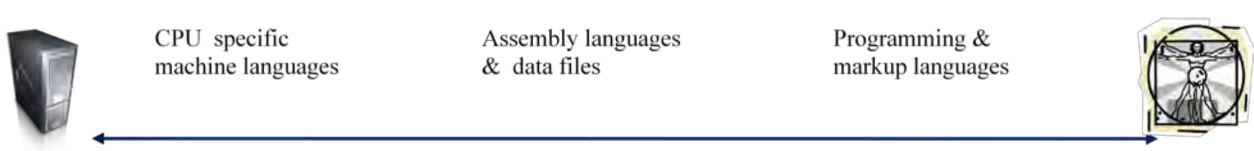

Figure 1: The range of intended audience for machine-oriented standards

With respect to machine-human interface standards, we note that there is a similar spectrum ranging from representations that are more machine-oriented to those that are more human-oriented. At the human end of the spectrum, controlled vocabularies and classified (bibliographic) indexing systems are often used to create the "external cognitive scaffolding" to aid humans in their interaction with the system in terms of search and retrieval (Jacob, 2004). The actual data within the system is typically organized according to logical structures such as ontologies, taxonomies, or schemas. While users typically do not interact with these structures directly, developers often describe or document them in non-machine interpretable forms, including written descriptions or pictorial representations. Finally on the more machine-oriented end of the spectrum there are file formats, often called mark-up languages, that are used to represent these structures in machine interpretable (but still human readable) form. These include XML, UML and OWL, among others.

Figure 2: The Range of Intended Audience for Interface Standards

We see a similar distinction among human-oriented standards. While human-oriented standards are, by definition, written for a human audience, some are closer to being parsable by a machine than others. The hallmarks of machine computable data are attribute defined classes, fixed semantics, and formalized syntax, and in so far as some human-oriented standards have these qualities, we can assert that they are more easily used in the context of machine computation. Categorizations that allow for overlapping categories and that do not explicitly define terms for the categories (e.g., reference books are one such example) are arguably the most purely human-oriented and the most distant in terms of their ability to be captured in an information system in a manner that captures its definitional structure. Dictionaries are based on categorizations (overlapping categories), but add additional structure by binding terms to the definitions, and in so doing, making the definitions (if not the concepts they represent) discrete from each other. While the existence of synonyms and hyponyms and the existence of definitions that conceptually overlap prevent machine reasoning for dictionaries, the discrete nature of the definitions along with the binding of labels, allows the content to be (mostly) captured in an information system such as Princeton University's Wordnet database (Princeton University, 2010). We would argue that standards systems that non-ambiguously bind single labels to single categories (or classes) are even closer to being machine processable, since all identified members are members of the same well defined set. These would include nominal identifier systems and ordered identifier systems, among others. In particular, we note that controlled vocabularies, taxonomies, and classified (bibliographic) indexing systems encode richer, more complex, relationships among the elements, and in this way, are even closer to forms traditionally used in computing. This is not surprising since we note that despite the fact that these human-oriented types all predate the use of computer systems, these three (controlled vocabularies, taxonomies and classified indexing systems) have traditionally been associated with the management of large corpuses; a usage that we argue presaged the use of automated information systems.

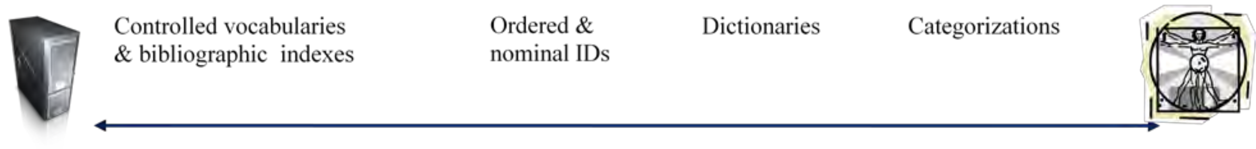

Figure 3: The Range of Intended Audience for Human-Oriented Standards

Summarizing to this point, we recognize three basic groupings of standards that range from machine-oriented to human-oriented: machine-oriented, human/machine interface, and human-oriented standards. And we observe that the types of standards within each of these groups themselves range from being more machine-oriented to being more human-oriented. To this we add that the three individual spectrums overlap. Humanly readable structured data file formats such as mark-up languages like HTML, XML, and OWL are the most human-oriented of the machine-oriented standards and at the same time, are the most machine-oriented of the human/machine interface standards. In like manner, controlled vocabularies and classified (bibliographic) indexes are among the most human-oriented of the human/machine interface standards and the most machine-oriented of the human-oriented standards. This leads us to assert that three spectrums can be combined into a single spectrum as shown in Figure 4.

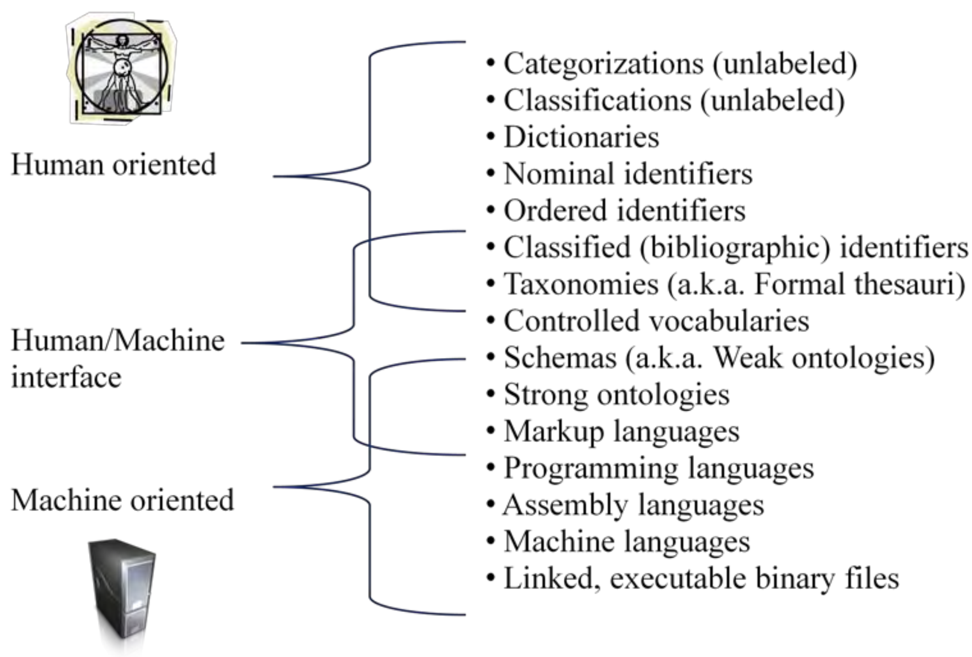

Figure 4: The Intended Audience Spectrum

2.2 Encoded Detail

The second perspective from which to view standards can be described in the amount of detail encoded in their definitions, which range from a little to a lot. Our assertion, which we hope to establish in this section, is that as we move across the spectrum of types, from human-oriented to machine-oriented, the amount of definitional detail encoded increases.

Our line of reasoning is largely based on Leo Obrst's concept of an "Ontology Spectrum", which he defines as "a description of the range of semantic models, as they increase in expressivity and complexity of structure" (Obrst, Forthcoming). Obrst defines this Ontology Spectrum for the set of standards that we refer to as interface standards, since they sit on the interface between machine-oriented standards and human-oriented standards (Mann and Brooks, 2010). Our argument seeks to extend Obrst's spectrum in two directions. First, we will argue that human-oriented standards have a lower amount of encoded detail while machine-oriented standards have more.

Second, while Obrst emphasizes complexity of structure among the categories (and classes) that are defined, we additionally distinguish between the complexity or precision of the semantics used to define a single category (or class). This allows us to distinguish between human-oriented standards in which a concept may be defined ambiguously and those which interface with machine-oriented standards that require more precisely defined concepts. We also distinguish among standards according to the amount of formality of syntactic expression. Summarizing, we define definitional detail within a standard (for classes, categories or members) as a composite of four aspects of definition including: the number of definitions, the complexity of the structural relationships among the definitions, the amount of semantics encoded within the individual definitions and the syntactic complexity of the definitional inscriptions and corresponding labels (when they exist).

We caveat our assertion as applying "in general" because there can be occasions when two types of standards have more or less definitional details of different kinds, making the comparison somewhat arbitrary. Never-the-less, we will argue that, on the whole, descriptive detail increases as standards become more machine-oriented.

Beginning with the most humanly oriented types, categorizations and classifications that do not a have labels associated with definitions have the least amount of definitional detail. In both forms, the definitional inscriptions may take the form of paragraphs or prose that adequately define the classes or categories, while the inscriptions themselves may not be explicitly distinguished from each other. Despite these similarities, we note that arguments can be made that either categorizations or classifications have the most detail, depending on which aspect of the composite of detail is discussed. Categorizations allow for arbitrary relationships between the categories and thus, structurally they form graphs which are more complex than the hierarchical (tree) structure of a classification. On the other hand, the class definitions in a classification must non-ambiguously define its classes in such a way that a) each class is a proper subset of a parent class and b) no two classes at the same level overlap. In this way, we would say that the definitions in a classification must capture more semantics than the definitions within a categorization, which may be more ambiguous and permitting of overlapping categories.

Dictionaries and glossaries add more structural detail in that they assign specific labels to each definition. While it is possible for a paragraph of prose to non-ambiguously define several terms and function as a dictionary, even though the definitional inscriptions are co-mingled, it is more common for dictionaries to discretely separate definitional inscriptions from each other (i.e. separate entries) and, in this way, they have more detailed definitional structure than many categorizations and classifications. While dictionaries associate terms with definitions, they do so in a way that is less structured than controlled vocabularies and identifier systems. The same term can be associated with multiple definitions (homonyms), and the same definition can be associated with multiple labels (synonyms).

The rest of the human-oriented standards all add additional definitional detail by imposing a single unique label for each definition. Nominal identifiers define a single category or class of objects, and for each member that is defined, add unambiguous semantic detail by associating the definition with a unique nominal identifier. Ordered identifiers add more detail in that the identified members must be ordered in some naturally recognizable way. Classified (or bibliographic) identifiers add further detail by enforcing an agreed upon classification of each identified object according to some separate but related classification system. While taxonomies do not necessarily create structured identifiers for all classified objects, they do add a further definitional detail by imposing a single term for each class within the related classification. Controlled vocabularies (or thesauri) are similar, but they impose yet further detail by enforcing agreement on a richer set of non-hierarchical relationships among the label/definition pairs.

Among the interface standards (not yet mentioned as overlapping with human-oriented standards), a schema (or weak ontology) is similar to a controlled vocabulary but adds enough additional semantics on the structure of the definitions and relationships so as to allow the information to be captured and parsed by an information processing system. (Strong) Ontologies add even more semantic detail so as to allow machine reasoning and inference based on the data (Obrst, Forthcoming)[4]. Schemas and (strong) ontologies can be rendered in non-machine readable form, such in paper renditions. For them to be captured in machine readable form, they are typically rendered digitally in some form of a markup language such as XML, UML, or OWL, and in so doing, additional syntactic detail is added.

Among the machine-oriented standards (not yet mentioned as overlapping with interface standards), so-called third generation programming languages are built upon their own schema or ontology that describe the constructs of the language and add to this the additional syntactic restriction of a formal grammar allowing programs written in the language to be machine parsed. From this point on through the spectrum, neither semantic nor syntactic detail increases but the sheer volume and complexity of the interrelationships increases. Programming languages are compiled into assembly code, which must add additional information that is particular to a set of CPUs. The standardized compilers thus need to understand the syntax of both the programming language and the syntax of the assembly language. Lastly, when installed and linked, a program is converted into machine code, which adds additional information about the local computer that the program is running on. At each step in the process, more detail and complexity is added.

2.3 Diversity of Practice

The third perspective from which to view standards can be described in the amount of diversity of practice among the groups that mutually affirm the standard, which ranges from a lot to a little. The assertion that we hope to establish in this section is that as we move across the spectrum of types, from human-oriented and less detailed to machine-oriented and more detailed, the diversity of the groups that can access and use the standard decreases.

By diversity, we primarily mean diversity of work practices. We ascribe to the theory that shared understanding of definitions of objects is created socially in the context of shared work practices that form mutual accountability among the participants. Groups of people recognize the same objects in the same way when they are members of the same (or similar) groups that do things

[4] As noted above, the increase in detail among interface standards was first observed by Obrst in his discussion of The Ontology Spectrum. See (Obrst, Forthcoming) for a more detailed discussion.

the same (or similar enough). (Bowker, et al., 1999) (Rawls, et al., Forthcoming) (Rawls, et al., 2009). For this reason, we prefer the term "community of practice" over "community of interest". There is a parallel understanding of diversity that applies to computers. Computer systems can be thought of as forming families based on similarities among CPUs and operating systems. In a very real way, different computers that share the same (or similar) CPUs and the same (or similar) operating systems literally do things the same way.

Unlabeled categorizations, such as reference books and standardized procedures, are among the most widely accessible type of standard. Object definitions form overlapping categories with imprecise boundaries, which allow for variance in judgments and interpretation among users. Further, no single label for definitions is enforced, allowing multiple communities of practice the freedom to refer to the same object definition using different terms. Unlabeled classifications constrain diversity of practice by enforcing the condition of proper subset inclusion of a parent/child relationship. In other words, unlabeled classifications can only be recognized by groups that agree on that particular classification structure. This is often not the case. For example, it has been shown that groups that agree on the basic level of a categorization may disagree at the super-ordinate (i.e., grouping) and sub-ordinate (i.e., sub class) levels (Green, et al., 2002). Dictionaries, which associate labels (i.e., words) with definitions, constrain diversity of practice in a different way. A group that associates a particular label with a definition can only recognize a dictionary as a standard if their label is included and correctly associated with its synonyms.

The rest of the human-oriented standards all impose the restriction of associating a single label for each definition. Nominal identifier systems require that groups that use the identifier system agree on the definition of the set that contains the list of identified members. This is frequently not the case as can be seen by the number of common objects that get assigned multiple nominal identifiers, which is rooted in disagreements among related communities on how to define the set of objects. For example, vehicles typically bear both license plate numbers and VINs. The former designates the car's membership in the set of legally registered vehicles while the latter designates its membership in the set of legally titleable vehicles. These are different definitions of set memberships that are used in different ways by different communities. In this way, we see that the use of nominal identifier restricts the diversity of the groups that can recognize the standard. Ordered identifiers and classified (bibliographic) identifiers similarly reduce the diversity of the groups that can recognize or utilize the standard by demanding that the users agree on the ordering or implied classification of the objects, respectively. The common use of classified (bibliographic) identifiers underscores how their additional structure limits diversity. For example, ISBNs, VINs, and Dewey Decimal Classification (DCC) numbers are typically read and used as nominal IDs, or at most, in the case of DCC numbers, as ordered IDs. However, in all three cases, experts in the respective fields can readily access the descriptive classificatory information encoded in the ID structure (Wikipedia, 2009b) (Wikipedia, 2009i) (Wikipedia, 2010).

Taxonomies and controlled vocabularies (thesauri) are even more restricted in their accessibility. For taxonomies, not only must the groups agree on a categorization scheme, they must also agree on single labels for all named classes (i.e., taxa) at all levels of the taxonomy. For controlled vocabularies (thesauri), users must additionally agree on a more complex set of relationships among the definitions. At each step along the spectrum, the need to agree on more definitional detail only serves to further constrain the diversity of the groups that can agree. This pattern continues among other interface standards.

Schemas and ontologies both provide more detail in order to make it possible for machines to parse the information and reason against it, respectively. In both cases, the additional detail demands that groups not only agree on the underlying controlled vocabularies, but also on the correct way to define and represent the classes for machine consumption.

When ontologies and schemas are rendered in truly machine readable forms, such as markup languages, the groups of machines that can utilize them become more limited. For instance, it is possible to render the same printed schema in multiple different markup languages. While these renderings may be logically equivalent for human readers in the written version from which they are derived, they are only accessible to those computer systems that have the proper software on them to read the different markup language formats.

In the same way, general purpose programming languages can only be compiled on systems that have the proper compilers. And once compiled, the resulting assembly language is further restricted to those machines with a CPU chipset that that recognizes that particular form of assembly. Lastly, once linked and installed, the resulting binary file will typically only run on the system that is was linked or installed on. At every turn, the addition of more detail limits the community of practice that can access the standard.

3 The Standards Trade-Space

In the previous section, we demonstrated that standards can be placed along a spectrum based on the intended audience which ranges from purely human-oriented to machine-oriented. Further, we showed that as we move along this spectrum towards machine-oriented standards, that the amount of definitional detail increases while the diversity among the communities of practice decreases. We summarize this in Figure 5.

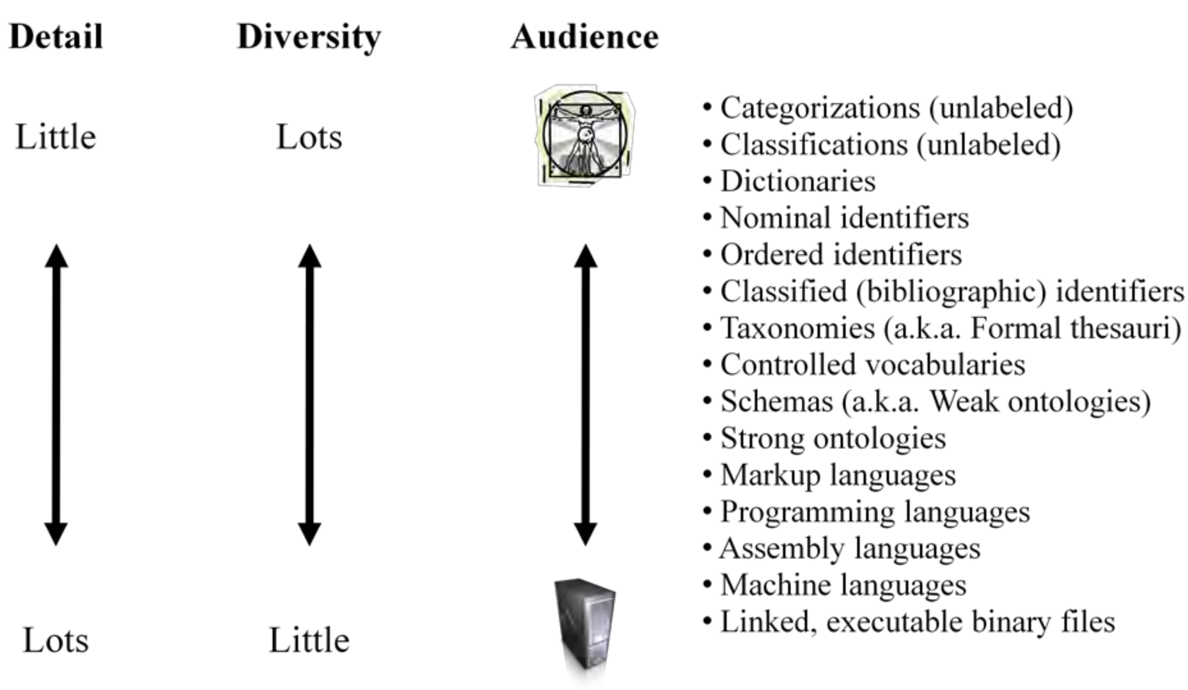

Figure 5: The Relationship between Detail and Diversity on the Intended Audience Spectrum

We are now in a position to make a fundamental observation regarding the relationship between standards:

> *The amount of definitional detail encoded in a standard is inversely related to the diversity of the community of practice (or groups of machines) that reference the standard.*

Based on this inverse relationship between detail and diversity among standards, we suggest that standards can be seen as existing along a trade space as shown in Figure 6[5].

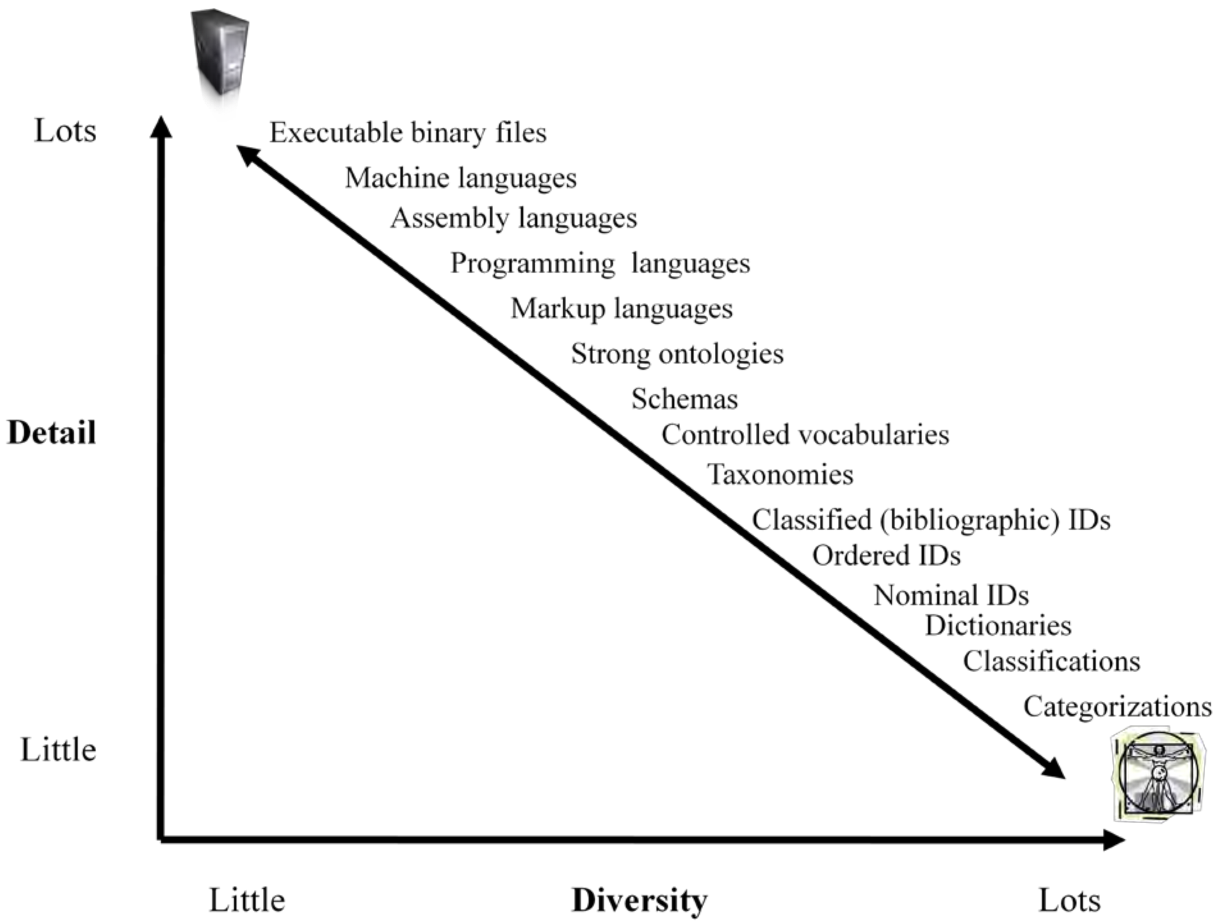

Figure 6: The Detail/Diversity Trade Space

We offer this trade space suggestively and recognize that additional research is required to determine the shape of the trade space curve and how much variance off of the curve is typical. Nevertheless, we assert that the fundamental inverse relationship on which the trade space is based is valid (as discussed in Section 2) and that the validity of the constraints imposed by the trade space is generally confirmed through experience in standards efforts.

Our experience with standards efforts leads us to make two further observations about standard efforts relative to the trade space.

[5] A trade space is a method of identifying optimal solutions in cases where variables are interdependent. They are also referred to as Pareto Fronts or Pareto Frontiers (Wikipedia, 2009j) (Wikipedia, 2010l) (Ross, et al., 2004).

3.1 The Zone of Inefficiency

Experience indicates that standards tend to not be found below and to the left of the trade space curve. We hypothesize that standards efforts tend to try to optimize along both primary dimensions: level of definitional detail and level of diversity of use. That is, if a standard is written at a particular level of detail, it will tend to be adopted by as broad a range of a communities of practice (or groups of computer systems) as can access that level of detail. Pictorially, we would say that standards tend to gravitate as far to the right as possible. On the other hand, if the diversity of usage across the communities of practice (or groups of computer systems) is fixed, we hypothesize that the broader community will aspire to codify as much detail as they can. Pictorially we would say that standards efforts tend to aspire to rise as far up as is possible. For this reason, we refer to the area below and to the left of the curve, the zone of inefficiency.

3.2 The Zone of Infeasibility

While experience indicates that standards efforts attempt to maximize both the level of definitional detail and diversity, experience also indicates that the inverse relationship between detail and diversity creates something of a limiting factor for standards. This experience indicates that standards development efforts that attempt to develop standards which provide too much detail for too diverse of a community tend to fail to be adopted. With respect to the trade space, we refer to the zone to the upper right as the zone of infeasibility.

Further, we can identify two typical paths that struggling standard efforts typically take that place them at risk to be in this zone of infeasibility. We refer to the first as "forced compliance." This takes place when a reasonably homogenous group achieves agreement on a fairly detailed standard and then attempts to demand compliance to that standard by a larger, more diverse set of groups. In these cases, the devil literally is in the level of detail. While the diverse group may agree on many general category definitions, the disagreement typically is found on matters of detail.

We refer to the second path that struggling standards can take as "death by committee". In this scenario, representatives from a diverse set of groups come together to attempt to agree on new shared standard with the aspiration that the shared standard will have the same amount of definitional detail that the individual groups enjoy in their own private, proprietary standards. Again, the problem is with the recognition of detail but this is different from groups disagreeing with an existing detailed standard that they are pressured/forced to utilize. Instead, the committee fails in their attempt to create a shared standard that has the level of detail that is expected.

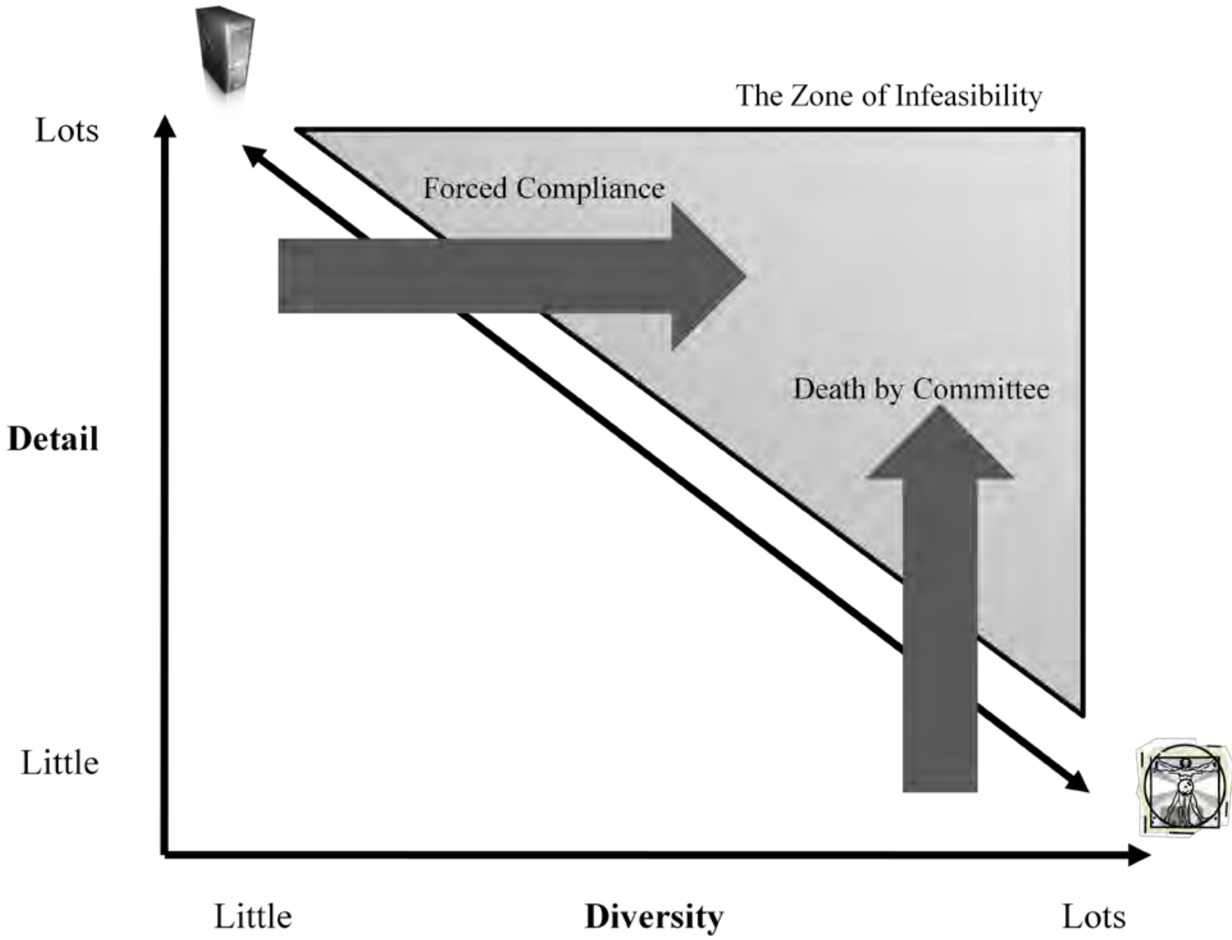

Figure 7: The Zone of Infeasibility

Both paths of failure are founded on false myths. The forced compliance approach is built on the false myth that others will be able to see things in "the same way we do." The death by committee approach is built on the false myth that diverse groups can collaborate to achieve the same degree of shared detail in common that they individually have on their own. Of course, these two myths are two sides of the same coin. There is an observed inverse relationship between the level of detail and the level of diversity that can be supported. When a high level of detail is achieved, there is a strict limit on the level of diversity of those groups that can utilize this detail. On the other hand, if a large degree of diversity is to be upheld, there is a strict limit on the amount of detail than be achieved.

4 The Evolution of Standards

In Section 2, we established that standards exist along a spectrum that ranges from human-oriented to machine-oriented. We also observed that the level of definitional detail encoded in standards and the amount of diversity of practice by groups that can recognize the standards also varies along this spectrum. In Section 3, we restated these relationships by describing them as forming a detail/diversity trade space and argued that standards tend to form along a trade space curve. In this section we consider some of the ways that standards can move or change over time.

Standards are typically associated with a feedback loop that exists in the analysis ecosystem in which a standards body takes as input a wide variety of information products produced in the field and produces as output an information product that is then used as a standard by analysts within the field. One direct implication of this is that the standard changes and evolves over time. Everyday examples of this include dictionaries such as the Oxford English Dictionary, which is updated on a regular basis (Wikipedia, 2010i). On the other hand, change is constrained and once a standard becomes deeply embedded, it becomes harder to change. So we have two influences to consider relative to the evolution of standards. One is their tendency to evolve over time. The other is their tendency to stabilize.[6] We consider these in turn.

4.1 Change

An example to consider is the evolution of the International Classification of Diseases (ICD), which is produced by the World Health Organization and which is detailed in "Sorting Things Out: Classification and Its Consequences" by Bowker and Star (Bowker, et al., 1999). The ICD traces its history back to the London Bills of Mortality, which was produced circa 1675. Bowker and Star discuss many of the sociological factors involved in the evolution of the classification of diseases, including the evolution of the organizations that were recognized as being authoritative. Summarizing the publication history:

- London Bills of Mortality - c1675
- Cullen's Classification - 1785
- International Statistical Congress - 1855, 1864, 1874, 1880, 1885
- International Statistical Institute - 1893, 1900, 1909, 1920, 1929, 1938
- World Health Organization - 1944, 1948, 1955, 1965, 1975, 1985

In so far as the published classifications reflect how disease was understood by the medical field at various times, this history reveals both an evolution of the classification and, just as importantly, an evolution of practices within the medical field as well. For example, Bowker and Star discuss how the disease of "hysteria" came to be recognized during the end of the 1800s and remained a recognized disease until the early 1900s. This is reflective of both changes in understandings of mental illness and different attitudes towards women. Attempting to bridge the gap between our concept of an analytical ecosystem and Bowker and Star's more general sociological point of view, we note that we must recognize that the analytical ecosystem of medical fields has always existed within the larger analytical ecosystem of society. So both perspectives have the basics of the feedback loop: the standards body produces the standard based on the current state of the field, and once published, the standard becomes one of many factors that influences the field to change.

The schools of thought influencing socio-technical studies have a variety of ways of understanding this feedback loop. Structurational socio-technologists would conceptualize the standard as a form of technology that co-evolves with the practices of its community of users (Barley, 1990) (Orlikowski, 1992) (DeSanctis, et al., 1994). Socio-technologists working from Communities of Practice and Ethnomethodology perspectives would emphasize that as a community is more unified in its practices relative to a set of objects, its members are able to recognize more shared meaning associated with those shared social objects, such as standards (Lave, et al., 1991) (Brown, et al., 1991) (Garfinkel, [1948] 2005). The difference between these schools of thought may be understood by considering the standard for measuring temperature. If

[6] The duality of stability and change figures prominently in the literature of contemporary organizational science and social theory. See e.g. Giddens (1984); Orlikowski (1992); Farjoun (2010).

one begins by emphasizing the evolution of the standard itself (i.e. from a structurational perspective), one might conclude that the standard is adapted over time according to the needs and practices of its users. Whereas ethnomethodological socio-technologists would emphasize that it is the shared practices of building and reading thermometers in similar ways that allows community members to conceive of temperature in shared or standardized ways. Despite these differences in emphasis, both viewpoints agree on the basic trajectory of this co-evolution between standards and their supporting communities of practitioners.

Bearing this in mind, we can now consider two common analytical goals within certain organizations of interest, namely a) to produce and access shared statistics and b) to automate the information processing required for generating and managing those data records necessary to produce statistics. Mutually recognizable statistics arise out of, among other things, commonly-developed human-oriented standards recognized by the various analytical processes (usage practices) within the organization. Or more precisely, the statistics are shaped in accord with the mutually recognized categories (or social objects) that are codified in the various human-oriented standards (Bowker, et al., 1999) (Rawls, et al., 2009). This happens in the context of the feedback loop that exists between the standards (including those that define the statistics) and the communities of practice that use, and are shaped by, those standards. The publication of the standards enables and constrains unity of practice, which in turn allows the communities to become more unified in their practices. Then the increase in unity of practice allows the groups to collectively recognize more shared detail which then gets encoded in the updated standards. In this way, the typical evolution is: a) the formation of human-oriented standards (and the categories they contain), b) the formation of statistics based on those standards (and their underlying categories) and then c) the automation of the data management that produces those statistics based on the standardized understandings of the categories, possibly followed by drift in practices away from common usage patterns to meet emergent (new) requirements and forcing the standards to change or suffer from breakdown (i.e. failure to remain "common"). We summarize this evolutionary process in Figure 9.

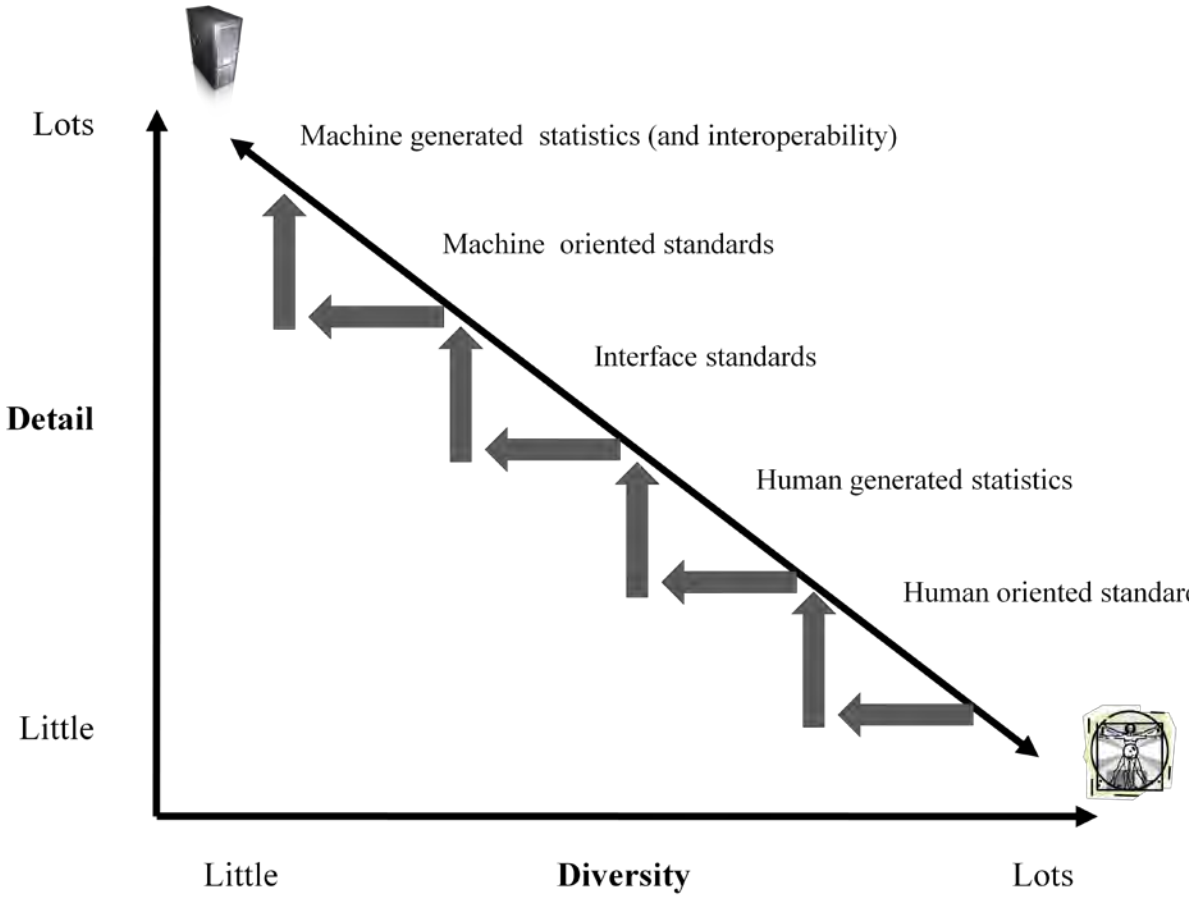

Figure 9: The Evolution of Standards, Statistics and Interoperability

This process works well when considering traditionally scoped information processing systems. So long as the user base is sufficiently unified as it often is within an organization, the system can be used by the users, and it will produce statistics they recognize. However, the situation becomes problematic when the information system is scaled to the enterprise level and asked to integrate statistic bearing data categories from multiple parts of the enterprise, or equivalently, with groups outside of the organization (Edwards, et al., 2007). Enterprises often do not have unified continuous leadership and are more often contested fields politically and culturally. Diverse groups have diverse practices, and as a result, they produce categories in different ways, and as a result, different statistics (Smith, 1974). The goal of achieving the same level of automation that is achievable with unified groups, with groups that are more diverse, places the development effort of the enterprise solution in the "zone of infeasibility". To see this, we run the chain of reasoning backwards along the evolutionary path as shown by the grey arrows in Figure 9. Namely, the statistic generating data records of the various sub-systems are irreconcilable in the enterprise system because the (pre-automation) statistics recognized by the various sub-groups are different. And, these (pre-automation) statistics are different because the human-oriented standards recognized by the analytical processes within the various sub-groups are different (Smith, 1974) (Garfinkel, [1948] 2005). That is, the information systems that are used by, and work well for, the various sub-groups are produced by evolutionary processes; and since the sub-group starts in different places, the information systems that evolve out of them end up in places that are different. More to the point, these differences remain irreconcilable if the reconciliation is attempted at the same level of detail with the goal of high levels of interoperability.

4.2 Stability

While it is possible for standards to evolve over time, it is also possible for them to both stabilize and to become essentially unchangeable. The process by which this happens is discussed in "Standards and their Stories" (Lampland, et al., 2009), where it is descriptively referred to as a "Russian doll effect".[7] Once a standard becomes established, it is available to be referenced by other standards. And once those standards become established, they too can be referenced by yet other standards, with each new layer of reference establishing another "doll". Eventually, the inner most standard becomes locked in, as any substantive change in that standard would require changes in all of the outer standards that rely on it either directly or indirectly.

The example discussed by Lampland, et al., in this regard is calendar age, which first became established as societal standard, following the American Civil War, as the U.S. government dealt with the large scale problem of dispensing veterans' benefits.[8] Once calendar age was established as a standard (or equivalently, date of birth), it became available for use for other standards that equate rights and responsibilities with age, generally on the assumption that age is a reliable correlator with either needs or abilities, such as: the right to drive, the responsibility to register for the selective service, or the right to drink alcohol. And once date of birth became an established standard on drivers' licenses (which themselves became established as standards), the age bearing licenses became available in laws requiring buyers of alcohol to be able to present documented proof of their age, often in the form of a driver's license. Layer upon layer of standardization has thus accrued around calendar age and date of birth, with the consequence that it is practically inconceivable to abandon or change it (Edwards, et al., 2007).

The adoption and deployment of an information system can have a similar effect within an organization. For example, if the information system is the primary means of tracking customers, then it is possible that particular attributes used to describe "customers" within that information system will become available to other standardized processes within the organization. This availability may be mediated by the user interface of the information system, reports generated by it, and/or by data input/output capabilities provided by the system. In this way, it is possible for internal organizational standards to accrue around the core standard of "customer" as defined by the information system. Once this has happened, this particular data standard of "customer" becomes "locked in" and any changes to the form of the "customer" record, including wholesale replacement of the information system, becomes a difficult and costly proposition, as it requires changes to all company standards that refer to it either directly or indirectly.

5 Recommendations for Information System Design

Up to this point, we have been exploring some of the relationships between human and machine-oriented standards. We next consider implications of these relationships on the design of traditional information systems and on the design of information systems that operate on the

[7] Other terms commonly employed in social and organizational science for the same kind of effect are "structural inertia" (Hannan & Freeman 1984; Gilbert 2005) and "path dependence"(Mahoney 2000; Garud & Karnoe 2001).

[8] As an aside, the establishment of calendar age as a standard, including the evolution of U.S. Census recording procedures, is also discussed and serves as yet another standard that evolved towards greater detailed precision while the practices of Census recorders became more unified (Edwards, et al., 2007).

enterprise scale. However, before proceeding, we summarize the main points about the relationship between human and machine-oriented standards:

- The identified types of standards fall on a spectrum, based on their intended audience, which ranges from human to machine.
- As we move from human-oriented to machine-oriented, the definitions encoded in them become more detailed, while the practices of the communities (or groups of computer systems) that can recognize and use them become more unified (or less diverse).
- The inverse relationship between detail encoded and diversity, along with our observed experience in developing standards, suggest that successful standards can be characterized as lying along a "detail/diversity trade space" curve, which functions as a constraint. That is, a standard that attempts to encode too much detail for a set of related groups whose practices are too diverse is at risk of failure, and in this way, is said to be in the "zone of infeasibility".
- Standards can co-evolve with their attending communities of practice, allowing the groups to recognize more detail as their practices become more unified. In some cases, this can allow groups to mutually recognize statistics (based on their shared standards) and to build interoperable information systems (built on shared, standardized data models) to automate the generation of those statistics.
- A standard can become socially "locked in" when other standards reference it. This effect also applies to standardized data objects that are encoded in information products used by an organization, making it difficult or expensive to change systems at a later time.

We now turn our attention to exploring the implications of these points.

5.1 Traditional Information Systems

We first consider the implications on traditional system design. Compared to enterprise scale information systems, traditional information systems are designed for a relatively unified set of communities of practice. We would consider most commercial software products that are designed to targeted niche markets and most custom applications developed and deployed by organizations to be in this category.

5.1.1 Data Model Development

One of the essential questions of information system design is how to conceive of and model the information that is to be processed. A point often made in the training of data modelers (and often missed) is the necessity for the data model to reflect the categories and concepts that are recognized by the user base. Surprisingly, the situation may be easiest when the data modeler is not at all familiar with the categories of the users. Since the data modeler has few preconceived ideas of the domain, they are forced to rely heavily on category definitions that are affirmed by users. The published, formalized, human-oriented standards used by the user group can provide excellent insight into how to codify categories in the data model in ways that are recognizable to the group. These standards may apply to an entire field or discipline, which are published externally from the organization, or may be localized, proprietary, standards that are published internally by the organization itself. Often, internally published proprietary standards occur as standardized forms (e.g., purchase request forms and invoices among others) or procedure manuals.

On the other hand, data modelers can encounter more problems capturing so-called "common sense" concepts such as "persons". A person can be one thing in one context and a different thing in another. But, the very commonness of the concept can lead the data modeler to wrongly assume that his or her own "common sense" understanding of the concept is aligned with the users' understanding of the concept. In fact, the data modeler's conception may be very much out of step with the localized understanding of the users (Goodwin, et al., 1996) (Rawls, et al., Forthcoming). In these situations, the human-oriented standards utilized by the users can provide an informative guide to the data modeler to prevent this problem. By utilizing these standards as a guide in modeling, not only for domain specific concepts but also for so-called "common sense" concepts, the data modeler can more effectively create models that are more recognizable and useful to the users. The issue is not what is "right" (according to the data modeler's common sense understanding), but what is "useful and recognizable" to the user group. If stable human-oriented standards exist and are used by the user group, and if those standards codify concepts in ways that are in conflict with the data modeler's understanding, then the standards affirmed by the user group should be used because the data modelers' versions will not be recognizable by the user.

Cases of IS design in which there are no stable human-oriented standards that are recognized by the user group can be even more difficult. In these cases, the data model and the information system that is built upon it may be the first occurrence of a codified standard. The information system and the reports generated from it act as a form of proxy for more traditional, written, human-oriented standards. In this situation, we offer two observations that have a bearing on the development of the information system. First, the fact that human-oriented standards co-evolve with the group suggests that it should be expected that the new information system, which is serving as a proxy for an otherwise non-existent human-oriented standard, will likewise co-evolve. That is, we should expect that systems built on more stabilized and established human-oriented standards will be able to use data models that are similarly more stable, while systems built in the absence of stable and established human-oriented standards will only be able to create data models that can be expected to be unstable and in need of comparatively more frequent updates. And indeed, this is why the rollout of a new system often needs numerous fixes before it sees wide usage.

More deeply, in situations in which there are no human-oriented standards that are affirmed by the user group, the data modeler should understand that the task that he/she is involved in is somewhat different than traditional data modeling. The data modeler needs to pay more attention to the sociological factors, such as the rights, privileges, and responsibilities that are mutually affirmed among the users, because these sociological factors are constitutive in the formation of a data model that captures categories that will be recognizable to the users (Rawls, et al., Forthcoming). In particular, for many data modelers this will mean less emphasis on the question of trying to define what the *thing* "is" (in a Platonic or realist sense) and more emphasis on the question of what people are trying to accomplish when they refer to the thing. For example, in the construction of an information system used to manage inventory for a car parts distributer, a traditional approach of attempting to define what a car part "is" led a team of data modelers to define parts according to the make, model, and year of the associated vehicle, along with classification information that described the kind of part such as drive train, brake, etc. However, this data model failed to account for attributes of parts that were critical to the task of inventory management - issues such as the part's dimensions, weight, and the length of time it has been in storage[9]. We take the ethnomethodological position that is shared social practices that create sharable categories (that can then be standardized in a data model). By focusing on the work

[9] Based on an interview with Dr. Joeseph DeRosa, of the MITRE Corporation

practices of the groups involved with inventory management, instead of relying on their own "common sense" understanding of parts, the data modelers would have been in a better position to determine which aspects of parts should have been standardized in the data model.

The philosophical position implied by traditional data modeling is the same as traditional understandings of category theory, which is that a thing or category can always be adequately described by a set of essential features that are inherent to the thing and thus, are universally meaningful (Rawls, et al., Forthcoming) (Murphy, 2002). It is this approach to data modeling that can lead to the creation of car part inventory systems that fails to account for size and weight. The more general point to be made here is that when the data modeler finds himself/herself in a situation where there are no stable and accepted human-oriented standards, she must recognize that she is no longer merely creating a data model for a computer system to execute against. Instead, she is involved in the necessarily sociological work of understanding how recognizable categories (i.e., social objects) are formed; work that demands that attention be paid to what the users are attempting to accomplish, and away from definitions of what the thing "is" in terms of essential features.

5.1.2 Specialized Software and Niche Markets

It is common place in the software development industry to distinguish between general purpose software that is designed to be used across a broad range of customer types and specialized software that is designed to be used by a more tightly defined niche market. The trade space and standard evolution and stabilization sections above give some insight into why this phenomenon is to be expected and provide a basis for making some general observations that may apply to software developers and organizations involved in the acquisition of such systems.

The basic distinction between special purpose software and general purpose software can be described directly by using the trade space perspective. Specialized software contains more detailed constructs that are germane to a relatively small and unified set of groups, whereas general purpose software can be used by a relatively larger and more diverse set of groups, although it lacks the high level of detail or functionality that might be useful for one particular group. Specialized software offers more functionality for a particular group; but it also has the reputation of being more expensive, since it typically has fewer members of the workforce that are competent in using the software, and it is therefore less commoditized. In contrast, general purpose software offers less functionality for any particular group, but typically has more users in the workforce who are competent in using it; the software therefore is more commoditized and develops the reputation of being less expensive

Understanding vertical markets and supply chains as information ecosystems of companies that exchange published information products may provide insight into how specialized software can come to be created. It is common to model vertical markets as a supply chain in which companies at one level of the supply chain produce goods and services that are consumed as resources by companies at the next level of the supply chain. Between each level, information products are exchanged, including lists of products and services, order forms, and invoices to name a few. In so far as the same kinds of supplies or products are required by all companies at a particular level of the supply chain, they competitively have access to the same set of suppliers (modulo locality constraints) and in this way, a feedback loop is created within the information ecosystem similar to that associated with formal standards. We may consider the relevant set of information products to constitute a loosely federated or non-formalized type of standard that then shapes the categories (through co-evolution) that are mutually recognizable among companies at both ends of this information product exchange. So, while it is the case that competitors at any particular level in the supply chain may have limited need or opportunities to engage in direct information exchange with each other, they may, nonetheless, have a lot of uniformity in their work practices and a lot of shared detail in the categories they are able to recognize.

We believe this is the context that provides the basis for specialized software products to be developed. The degree of similarity in work practices and the relatively high degree of stable concepts that are shared across the competitors at a given level within a vertical supply chain implies that a software product that aligns with common work place practice and that encodes categories recognizable among this set of competitors has the potential of providing levels of functionality that more general software products cannot.

With these observations, we can then crudely categorize traditional software products into three major categories along the trade space spectrum. General purpose software encodes widely understood concepts at low levels of detail, and are usable by a wide range of customers while providing comparatively lower levels of capability. Specialized software encodes concepts that are understood within a more tightly constrained market niche and does so at comparatively higher levels of detail which are usable by a more restricted range of customers while providing comparatively higher levels of capability. And, customized software encodes concepts that are recognizable only within the context of a single organization and does so at a very high level of detail and are usable by only a single organization while providing the highest level of capability.

We believe this insight has several implications for both software producers and consumers. For software producers, this suggests a choice between producing specialized, software or more general software solutions. We believe that many software vendors have already internalized this basic decision, and we offer little on this point beyond the observation that the choice is real. The more specialized they make their product, the more value it will bring to certain customers but the number of customers who will be interested will go down. In this way, the trade space may provide a mechanism that helps them to structure their market analysis and revenue forecasting.

For organizational buyers of software, the trade space and the resulting conclusions imply that there are meaningful pros and cons to be considered between general purpose, specialized and custom software. The more specialized software promises greater functionality, but only if the assumptions used in its creation meet the need. While the potential for real productivity gains from more specialized software has appeal, there are potential negative implications. The more specialized the software is, the smaller the set of potential competent users will be in the work force, possibly demanding higher salaries. Lastly, more specialized software solutions that utilize the software vendor's own proprietary data models create a form of standard within the organization deploying that software. That is, in so far as those data structures are encoded in the information reports generated by that software, the organization will layer other reports on top of those and the Russian doll affect will embed that software deeply in the organization. This, in turn, can have the effect of making the possible transition to another software vendor prohibitively expensive.

These tradeoffs may be partially complicated or relieved in cases in which the organization's field or discipline adopts more standards. In such cases, it is possible for an industry standard to materialize, become adopted by a broad range of vendors, and effectively supplant proprietary constructs. One of the appeals of software products that incorporate industry standard data constructs is that the data standard can serve to separate the software from the organization's embedded standards that rely on that data. That is, if the data coming out of an information system adheres to an industry standard, it is likely that replacement software that also adheres to that standard could be deployed with less cost. Organizations involved in the purchase and deployment of enterprise solutions might consider the potential of the development of human-level and interface level standards (that often deal with data format and transport) and consider transition strategies to products that incorporate those standards.

While this observation may imply that value is created for customers when software incorporates human or interface standards, there are also potential benefits for software vendors . For example, when customers demand the ability to integrate multiple products into a larger, more comprehensive solution, there is a benefit to software vendors who can effectively leverage shared information standards so that their products could more easily be integrated with other systems also adhering to the standard.

5.2 Enterprise Scale Information Systems

Despite the challenges mentioned above, the ubiquitous deployment and use of traditional information systems is evidence that traditional information systems can be built on data models that adequately capture categories and generate statistics that are recognizable to their user base. However, enterprise scale information systems are often less successful (Brooks, et al., Forthcoming). One of the many challenges in engineering enterprise scale information system is the creation of data models that have meaning across a diverse set of groups while at the same time achieving the same level of detailed automation associated with more narrowly focused information systems. Stating this in terms of the detail-diversity trade space, the stated goal of some enterprise scale development efforts places them in the "zone of infeasibility" by attempting to encode too much detail for too diverse of a population.

We believe that the constraints of the detail-diversity trade space, the distinction and relationship between human and machine standards, the evolutionary development and solidification of standards all provide a basis for making recommendations for developing and managing information systems in an enterprise scale context.

5.2.1 Avoid "Forced Compliance"

The error of "forced compliance" occurs when a relatively unified group has produced (or co-evolved with) a highly detailed standard (such as an information system, data model or a traditional human-oriented standard) and attempts to mandate that other groups adopt or use that standard. This approach fails because different groups with different practices conceive of concepts differently and as a result, the proposed standard fails to be useful for the different groups. When the group proposing the adoption of the highly detailed standard lacks the political or economic power to enforce compliance, the different groups have the option of simply ignoring both the request and the standard.

When the group proposing the standard does have the political and economic power to compel compliance, the results can produce poor data and unreliable statistics. The groups being forced to comply with the standard differ from groups that created the standard because they have different practices. These differences in practices render the categories codified in the mandated standards as unrecognizable or subject to large degrees of variance in terms of interpreted applicability (Smith, 1974). In short, the group demanding compliance is demanding something of the other groups that they do not have the wherewithal to provide. Since they do things differently, they cannot recognize the same categories, and because of this, any data or statistics submitted in the name of compliance cannot be relied upon as having the intended meaning (Garfinkel, [1948] 2005).

When looking to establish linkages to an existing highly detailed information system, the constraints of the detail-diversity trade space imply the possibility of two alternatives to the "forced compliance" march into the zone of infeasibility. The first is to limit the set of participating groups to only those who have enough uniformity in practice to be able to jointly recognize the full detail of the information system (and the standardized data model on which it is built). The other is to decrease the amount of detail expected in order to gain broader acceptance. Loose couplers are simplified data sets and their associated services that can have functional meaning across a broader set of groups within the enterprise. The shared data models and supported functionality are limited in comparison to more fully realized localized information systems, but they are usable across a broader set of users as indicated in the detail-diversity trade space (Miller, et al., 2007). The incorporation of even less detailed human-oriented standards such as classified identifiers (e.g., VINs) or nominal identifiers (e.g. license plates) can allow information products and data records to be more quickly correlated by human analysts that use the information system. In this way, they operate as realized "boundary objects", allowing different groups at the enterprise level to coordinate their work with each other more effectively (Star, et al., 1989) (Lee, 2005) (Rawls, et al., 2009).

5.2.2 Avoid "Death by Committee"

The error of "death by committee" occurs when representatives from multiple diverse groups are brought together and tasked with creating a new information system (or data model or human-oriented standard) with the expectation that they will be able to agree on the same high level of detail associated with their individual native information systems. In situations in which the groups have significantly different native practices, the groups, will conceive of related concepts differently. This tends to lead to protracted disagreements on how to construct a single comprehensive data model (or system or standard), a situation sometimes referred to by data modelers as "schema wars" (Rawls, et al., Forthcoming).

Again, when starting with a collection of diverse groups that desire to forge agreement on detail to support interoperability among their systems, the constraints of the detail-diversity trade space offer two alternatives. First, the collective set of groups can lower their expectations in terms of the amount of detail they can aspire to agree upon. Again, lower detailed solutions like loose couplers might be considered. However, with highly diverse groups even this degree of detail may be too much, in which case seeking agreement on less detailed human-oriented standards may be the next best possible achievable outcome.

Alternatively, if it is mandatory that some groups achieve agreement on a lot of detail, the number of participating groups can be reduced to help reduce the amount of diversity among the groups and thereby increase the amount of mutually recognizable detail. Again we emphasize that it is shared practice that allows for shared concepts. Thus, identifying groups that have similar practices will maximize the opportunity for the involved groups to achieve agreement.

5.2.3 Invest In Human-Oriented Standards

The situation in which the enterprise or field desires to have shared statistics and increased automation but lacks stable and accepted human-oriented standards can be particularly daunting. The constraints of the detail-diversity trade space suggest that both goals will be unrealized since the groups are too diverse to recognize the shared categories necessary to construct the statistics and to automate.

In this situation, the idea of co-evolution of groups and their standards offers some hope. Provided that a) the underlying practices of the groups are similar enough to each other and b) the groups have a shared goal or mission that necessitates that they collaborate effectively, it may be possible to create human-oriented standards that are mutually recognizable for all of the groups involved. Once the human-oriented standards are established and stabilized, then it may be possible to create mutually recognizable statistics and automated information sharing and processing based on the shared categories codified in the standards.

We should make several points about this approach. First, this approach recognizes that the success of all information systems (and statistics) rests on a foundation of stable social practices that are shared among the users. This approach understands that the lack of shared human-oriented standards is a risk factor, since it indicates that the social foundation for acceptance doesn't exist yet. Moreover, just as a farmer prepares the soil before planting the seed, by directing investments toward the establishment of the missing human-oriented standards, this approach seeks to influence the analytical ecosystem and to create a situation in which the construction of shared statistics and their supporting information systems might succeed.

Second, this approach recognizes that the systems engineering work to be done is outside of the realm of traditional systems engineering. The focus shifts away from the formalization of statistics and metrics, the establishment of shared data schemas and the engineering of information systems; rather it shifts towards building cross-organizational consensus on standardized practices, concepts and terms, and the establishment of an on-going analytical capability to produce and maintain the standard. This on-going analytical capability plays the role of a standards body and works to publish and drive adoption of the new standard (thereby closing the analytical feedback loop typically associated with human-oriented standards). For example, the Lean Aerospace Initiative (a consortium of major aerospace companies and suppliers, government agencies, organized labor and researchers from MIT's Engineering Systems Division) found that developing a long-term community of practice, which could agree on next steps, required recurrently meeting in a neutral forum, generating a common knowledge base, and developing a common vocabulary, all within the presence of a trusted change agent (Murman et al 2002).

Third, and related closely to the second point, since the work of establishing new human-oriented standards is fundamentally social in nature, this approach recognizes that the role of the designer of a human-oriented standards is one of influence, not control. This is reflected in the tension between the editorial postures of a) prescription, which aspires to enforce right use of terms and concepts, and b) description, which aspires to capture and reflect current usage (Mann and Brooks, 2010). While noting the distinction in these approaches, we recognize that even the most prescriptive editorial posture is constrained by acceptance within the field, or more directly, by its lack of acceptance. That is, if the designer of a human-oriented standard prescribes concepts that are entirely out of line with current practices, those concepts are at risk of not being comprehensible to the groups and thus, likely to be rejected in subsequent revisions of the standard. We point to Bowker and Star's example of the removal of "hysteria" from the International Classification of Disease as a case in which changing practices render a category meaningless (Bowker, et al., 1999).

The folklore of astute campus planners is that it is better to postpone the paving of sidewalks between campus buildings and wait until the common usage of foot traffic impacts the grass enough to determine where sidewalks will be accepted and used. In the same way, even in the most prescriptive of circumstances, human standards flow out of, and are constrained by, the established patterns of practices within the groups. Engineering practice and philosophy presume that the usage patterns are known, fixed, and codified in stable requirements statements, and the goal of engineering is to enforce control to ensure that the developed system reliably and repeatedly produces consistent results. In contrast, the establishment of a new human-oriented standard is better understood as design. While there is a degree of influence that can be brought to bear through a more prescriptive editorial posture, the standard designer cannot control the outcome in the same way that engineers have traditionally. Ultimately, the practices of groups are likely to step off of the proverbial prescribed sidewalks and for the standard to remain relevant, the standard must evolve.

Fourth, and again closely related to the second, this approach recognizes and accepts that creation and management of human-oriented standards is generally very different in terms of programmatic realities than the development and deployment of machine-oriented standards. Machine-oriented standards tend to be published and revised as versioned releases. Typically, there is an increase in the level of development effort leading up to the release of a new version followed by a relatively low period of maintenance that continues until a new update is undertaken. Eventually, as the machine-oriented technical standard stabilizes the periods between updates grows longer. Many examples of this can be seen among commonly encountered internet standards such as the Internet Protocol (IP), the Simple Mail Transport Protocol (SMTP), and the Hyper-Text Markup Language (HTML).

In contrast, while human-oriented standards stabilize, it is also true that they typically get updated more incrementally on an on-going basis. Dictionaries, encyclopedias, and even the International Classification of Diseases all get updated on a regular on-going basis, even if it is years between updates (Bowker, et al., 1999) (Wikipedia, 2009). Additionally, other human-oriented standards require on-going human analysis to produce on an on-going stream of new content within that standard. For example, while the Library of Congress classification system may continue to evolve slowly over time, the publication of new books is on-going, requiring a similarly on-going task of analyzing new books and assigning them their official LOC classification and ID (Wikipedia, 2009f). Similarly, U.S. Social Security Numbers need to be assigned on an on-going basis (United States Social Security Administration, 2010). The implication here is that the level of effort and funding levels associated with human-oriented standards tend to ramp up to their initial development, and then they often maintain a substantial level that needs to be sustained on an on-going basis.

Fifth, this approach of investing in human-oriented standards with the hope of creating an environment that will support the future development of enterprise scalable statistics and information systems must be recognized as a long term strategy. The stabilization of human-oriented standards can take years or even decades to occur. And as the example of the International Classification of Diseases reveals, the co-evolution of groups and their human-oriented standards can continue for centuries.

But, we point to the 10 year history of a case in which the introduction of a new human-oriented standard eventually led a field to recognize, codify and automate sharable statistics - that of MITRE's Common Vulnerability and Exposure's list (CVE) - as a more rapid example[10]. Prior to 1998, the information security field recognized the concept of a software vulnerability as a flaw in a computer program or operating system that would allow a malicious attacker to compromise

[10] We believe that due to its relative immaturity as a discipline, that the field of information security provides a particularly interesting context to consider the development of standards. The practices within information security are comparatively young and in many cases, are still forming. Correspondingly, the standards that are recognizable within the field are in their early stages of development and have less "Russian Doll" effects in place.

the system and to gain unauthorized access to data or functionality. However, there was no agreed upon way to either objectify or count vulnerabilities and as a result, there was no way to efficiently correlate vulnerability information across different information sources, tools and services. In 1998, the MITRE Corporation launched the CVE program which created an enumerative identifier system that a) discriminated among publicly known vulnerabilities based on input and feedback from the vulnerability management community and b) assigned nominal identifiers to items on the list (similar to license plate numbers), and c) counted them. Within 3 years, CVE IDs were used in most major vulnerability management products, security bulletins and other sources of vulnerability information. For example, in 2001 Microsoft began to include CVE IDs in their security advisories.

The trend continued in 2005 when the Forum of Incident Response and Security Teams (FIRST), which is the international coordinator among computer emergency response organizations, introduced the Common Vulnerability Scoring System (CVSS), which defined a standardized metric for risk associated with vulnerabilities (Forum of Incident Response and Security Teams, 2009). Subsequent to that, the U.S. National Vulnerability Database (NVD, managed by the US National Institute of Standards and Technology) began to publish CVSS scores for all vulnerabilities published in MITRE's CVE list. And after that, the VISA Payment Card Industry consortium instituted financial penalties against participating organizations for each vulnerability found on the organization's computers that scored above a certain threshold, as determined by NIST's published CVSS score for the related CVE. By itself, the CVE list is just a human-oriented standard, much in the same way as license plates and SSNs. It took several years before the use of CVE IDs became pervasive and more deeply used (for basic counting principles encoded in the list) to co-evolve with the vulnerability management community. But once CVE had stabilized and became broadly accepted, shared statistics could be defined, instantiated, published, and automated. It has taken 10 years, but the initial investment in the human-oriented standard (CVE) has shaped its supporting industry and allowed the goals of statistics and automation of processes to be achieved.

5.2.4 Achieving Immediate Interoperability

Investing in human-oriented standards as a long term strategy for creating the environment in which enterprise scale information sharing, statistics, and automation might be achieved is not always feasible. There are times where an automated capability must be deployed as soon as possible. In these cases, the detail-diversity trade space also offers some guidance on what can be achieved.

Interoperability demands a lot of agreement on detail and correspondingly tolerates little to no diversity. The direct implication here is that if achieving interoperability of systems is a non-negotiable goal, then the strategic focus must be shifted to managing and limiting the amount of diversity among participating organizations. We offer several options to consider. First, an organization seeking to achieve interoperability with other organizations should make an assessment of their political, economic, and operations power relative to other participants. The detail-diversity trade space predicts -- and experience with managing standards affirms -- that conflict and demands for compromise ensue when diversity of groups is introduced while, at the same time, the high level of detail needed for interoperability is demanded. Since agreement on a lot of detail is needed for full interoperability, either all organizations will need to make some adjustments to their existing systems (i.e., compromise) or a single organization's view will be adopted by all participants (i.e., forced compliance). The sociological understandings on which the detail-diversity trade space is based tell us that the differing sets of categories in use that lead to such conflicts among the groups is arbitrary in the sense that, for each group, the categories make sense relative to their own practices. The implication here is that there can be no single, technically, correct solution that will convince all others to change. So, as a result, decisions on detail will be made within the context of varying amounts of political, economic, and operational power.

Second and relative to the first point on power, an organization seeking to achieve interoperability with other groups should consider taking steps to exclude the participation of groups that cannot or are unwilling to adopt the shared standards needed to achieve the mandated interoperability. The detail-diversity trade space guidance on this point is that it is simply infeasible for multiple diverse organizations to achieve agreement on the high levels of detail needed for interoperability. Putting this another way, it is not possible for a wide range of groups to both a) each operate in their own unique and particular ways and b) interoperate with each other. The implication here is that an organization should restrict its interoperability goals to those organizations with closely aligned practices, while concurrently letting go of the goal to interoperate with groups whose practices are different. In a study of enterprise systems engineering projects, Brooks, Carroll and Beard (forthcoming) found that only those cases of enterprise systems engineering where the relationships between major stakeholders were already worked out could be considered "successful"; while those cases where technology was being used as the "tip of the spear" to force consolidation were unsuccessful. We therefore might call this posture the "Live, and let live" approach, or more correctly, "Operate, and let operate."

Third, an organization aspiring to achieve interoperability with other organizations should expect unreliable statistics to be produced by organizations at the periphery of the circle of agreement on the shared details of standards. It has been noted that recognition of a category is tied to group membership and that the recognition is not an all or nothing thing. Somebody who is considered to be merely competent relative to a group might recognize a category while for an expert that same category becomes fully naturalized (Bowker, et al., 1999). When closely related groups are combined, experts in one group might be considered to be merely competent in the other. In this way, categories necessary for shared statistics and interoperability might be naturalized by some of the participants and merely (or even barely) recognizable by others. This difference in group membership is related to the difficulties in producing workplace records as noted by Harold Garfinkel's "Good Reasons for Bad Organizational Records" (Garfinkel, [1948] 2005). Understanding the diversity axis of the detail-diversity trade space as diversity in usage practices gives both an explanation as to why this happens and a possible course to correct it. Variations on workplace practices is a key reason why statistics created in related but different groups may end up differently. Realities encountered in the workplace are equivocal, and how one group interprets and categorizes them may differ from another (Weick 1979; Brown & Duguid 1991). It is not that other groups will not choose to comply, it is that their differing practices make it impossible for them to comply (Garfinkel, [1948] 2005). A possible solution is to introduce human-oriented standards that focus not on the meaning of categories but on shared practices.

Fourth, if a group of organizations successfully manages to agree to interoperate with each other, they should anticipate the possibility of significant retooling costs in the future. Joint commitment by the groups to interoperate admits the possibility or likelihood of the standards, statistics, and information systems deployed to become embedded in the participating organizations due to the Russian doll affect. Concurrently, standards competing in spaces external to the organizations may evolve following a longer term evolutionary path that starts with broadly accepted but comparatively less detailed human-oriented standards. If those broadly accepted human-oriented standards stabilize sufficiently, broadly accepted categories and statistics can emerge, and with them, ultimately, broadly shared information structures. When this happens, the organizations that committed to more immediate interoperability with a limited number of partners based on proprietary standards will need to weigh the costs and benefits of continuing to use a firmly embedded set of standards with limited adoption against the conversion costs associated with adopting new, broad-based standards.

6 Conclusion and Future Work

In this paper we have explored relationships between human-oriented standards and machine-oriented standards. The first set of relationships we explored are the three spectrums of distinction among the standards, namely: a) the intended audiences, which range from humans on one extreme to machines (central processing units) on the other, with interface standards like schemas and ontologies in the middle; b) the amount of detail codified in the standard, ranging from (relatively) low for human-oriented standards to high for machine-oriented standards and; c) the amount of diversity among the communities (and groups of computer systems) able to access and utilize the standards, ranging from low for machine-oriented standards to high for human-oriented standards.

The second type of relationship that we have asserted is the trade space between the levels of detail and the diversity of the groups. That is, standards that aspire to exceed the constraint of this tradeoff by providing too much detail for too diverse of an audience tend to fail and enter the "zone of infeasibility".

The third and final set of relationships that we explored is the two interrelated forms of change that affect standards. The first is the process of co-evolution that allows standards to develop and the practices of the groups to become more unified. The second is the process of stabilization of standards that occurs as other standards are created in reference to the first, creating what has been referred to as the Russian doll affect that tends to "lock" standards in and making them difficult to change or replace.

In light of these relationships, we have offered several general recommendations regarding both traditional information systems and enterprise systems development. While we believe that the general thrust of our recommendations is sound, we invite further study to both validate and extend them. We would like to highlight one area in particular that we believe merits further investigation.

We discussed some of the mechanisms by which "niche markets" develop. A niche market occurs when a sizable market niche is serviced by a relatively small number of software manufacturers who produce specialized software that is targeted specifically to that niche while, at the same time, other software vendors produce more generalized software. While an organization may be able to get more functionality by using specialized software, it also runs the risk of standardizing its operations based on the proprietary internal data structures of the specialized software while the external environment continues to develop and evolve human-oriented standards that could eventually require the organization to replace that proprietary software with more standards based solutions. So it is this choice between specialized solutions, which tend to become deeply embedded in an organization, and more generalized solutions, which are more broadly accepted, which present difficult decisions in terms of systems development and deployment strategy. We posit this is an area that could be further illuminated by conducting case studies based on historical reviews of past development and deployment efforts that considered different deployment and transition strategies that have been used in the past from the view point of the detail-diversity trade space and standards evolution concepts presented in this paper.

7 Bibliography

Barley Stephen R. The Alignment of Technology and Structure through Roles and Networks [Journal] // Administrative Science Quarterly. - [s.l.] : Johnson Graduate School of Management, Cornell University, March 1990. - Vols. Vol. 35, No. 1, Special Issue: Technology,. - pp. 61-103.

Bowker Geoffrey C. and Star Susan Leigh Sorting Things Out: Classification and Its Consequences [Book]. - [s.l.] : The MIT Press, 1999.

Brooks JoAnn M., Carroll John S. and Beard Jon W. Dueling Stakeholders and Dual-Hatted Systems Engineers: Engineering Challenges, Capabilities and Skills in Government Infrastructure Technology Projects [Report]. - [s.l.] : The MITRE Corporation, Forthcoming.

Brown J.S. and Duguid P. Organizational Learning and Communities-of-Practice: Toward a Unified View of Working, Learning, and Innovation [Journal] // Organization Science. - 1991. - No. 1 : Vol. Vol. 2. - pp. 40-57.

DeSanctis Gerardine and Poole Marshal Scott Capturing the Complexity in Advanced Technology Use: Adaptive Structuration Theory [Journal]. - [s.l.] : INFORMS, May 1994. - Vol. 5, No. 2 : Vol. Organization Science. - pp. 121-147.

Edwards P. N. [et al.] Understanding infrastructure: Dynamics, tensions, and design [Report]. - [s.l.] : University of Michigan, 2007.

Forum of Incident Response and Security Teams Introduction to CVSS [Online] // FIRST. - FIRST.org, Inc., 2009. - March 2010. - http://www.first.org/cvss/intro/.

Garfinkel Harold Seeing Sociologically: The Routine Grounds of Social Action [Book] / ed. Rawls Anne Warfield. - Boulder : Paradigm Publishers, [1948] 2005.

Goodwin Charles and Goodwin Marjorie Harness Seeing as a Situated Activity: Formulating Planes [Journal] / ed. Engeström Y. and Middleton D.. - Cambridge : Cambridge University, 1996. - Vol. Cognition and Communication at Work. - pp. 61-95.

Green Rebecca, Bean Carol A. and Hudon Michele Universality and Basic Level Concepts [Conference] // 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, 2002.

International Commission on Zoological Nomenclature Introduction [Online] // INTERNATIONAL CODE OF ZOOLOGICAL NOMENCLATURE online. - International Commission on Zoological Nomenclature, May 1999. - January 2010. - http://www.iczn.org/iczn/index.jsp.

Jacob Elin K. Classification and Categorization: A Difference that Makes a Difference [Journal]. - [s.l.] : The Board of Trustees, University of Illinois, 2004. - Vol. 52, No. 3, Winter 2004 : Vol. LIBRARY TRENDS. - pp. 515-540.

Lampland Martha [et al.] Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life [Book] / ed. Lampland Martha and Star Susan Leigh. - [s.l.] : Cornell University Press, 2009.

Lave Jean and Wegner Etienne Situated Learning. Legitimate peripheral participation [Book]. - Cambridge : University of Cambridge Press, 1991.

Lee Charlotte P. Between Chaos and Routine: Boudary Negotiating Artfiacts in Collaboration [Conference] // ECSCW 2005: Proceedings of the Ninth European Conference on Computer-Supported Cooperative Work. - Paris, France : Springer, 2005.

Mann David and Brooks JoAnn Information Standards and Their Use: Implications for Design Patterns [Report] : Case Number: 10-1789. - [s.l.] : The MITRE Corp, 2010.
Miller Robert W. and Winkowski Daniel G. LOOSE COUPLERS AS AN INFORMATION DESIGN STRATEGY [Conference] // Military Communications Conference (MILCOM). - [s.l.] : IEEE, 2007. - pp. 1-6. - MILCOM.2007.4454906.
Murphy Gregory L. The Big Book of Concepts [Book]. - [s.l.] : The MIT Press, 2002.
Obrst Leo Ontological Architectures [Journal]. - [s.l.] : Springer, Forthcoming. - Vols. Chapter 2 in Part One: Ontology as Technology in the book: TAO – Theory and Applications of Ontology,Volume 2: The Information-science Stance.
Orlikowski Wanda J. The Duality of Technology: Rethinking the Concept of Technology in Organizations [Journal] = Focused Issue: Management of Technology // Organization Science. - [s.l.] : INFORMS, August 1992. - 3 : Vol. 3. - pp. 398-427.
Princeton University WordNet [Online] // WordNet. - Princeton University, January 2010. - February 2010. - http://wordnet.princeton.edu/.
Rawls Anne Warfield [et al.] Simple Enumerations. Ethnomethodology and MITRE Information Assurance Data Standards [Journal]. - [s.l.] : Il Mulino, 2009. - Issue 1 : Vol. Etnografia e ricerca qualitativa.
Rawls Anne Warfield and Mann David "The Thing is … What is Our 'What'?":An EM/CA Study of a Discussion of "Object" Clarity as a Problem in Information Systems [Report]. - [s.l.] : The MITRE Corporation, Forthcoming.
Ross Adam M. [et al.] Multi-Attribute Tradespace Exploration as Front End for Effective Space System Design [Journal] // Journal of Spacecraft and Rockets. - [s.l.] : American Institute of Aeronautics and Astronautics, Inc., January-February 2004. - Vols. Vol. 41, No. 1. - pp. 20-28.
Smith Dorothy E. The Social Construction of Documentary Reality [Journal]. - 1974. - (4) : Vol. Sociological Inquiry. - pp. 257-268.
Star S. L. and Griesemer J.R. Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professional in Berkeley's Museum of Vertebrate Zoology [Journal] // Social Studies of Science. - 1989. - Vol. 19. - pp. 387-420.
United States Social Security Administration Social Security Number & Card [Online] // Social Security Online. - USA.gov, January 2010. - March 2010. - http://www.ssa.gov/ssnumber/.
Wikipedia Dewey Decimal Classification [Online] // Wikipedia. - Wikimedia Foundation, Inc, August 2009b. - September 2009. - http://en.wikipedia.org/wiki/Dewey_Decimal_Classification.html.
Wikipedia Dictionary [Online] // Wikipedia. - Wikimedia Foundation, Inc., August 2009. - September 2009. - http://en.wikipedia.org/wiki/Dictionary.html.
Wikipedia International Standard Book Number [Online] // Wikipedia. - Wikimedia Foundation, Inc., March 2010k. - April 2010. - http://en.wikipedia.org/wiki/ISBN.
Wikipedia Library of Congress Classification [Online] // Wikipedia. - Wikimedia Foundation, Inc., August 2009f. - September 2009. - http://en.wikipedia.org/wiki/Library_of_Congress_classification.html.
Wikipedia Oxford English Dictionary [Online] // Wikipedia. - Wikimedia Foundation, Inc., February 2010i. - February 2/1/2010, 2010. - http://en.wikipedia.org/wiki/Oed.
Wikipedia Pareto Efficiency [Online] // Wikipedia. - Wikimedia Foundation, Inc., March 2010l. - April 2010. - http://en.wikipedia.org/wiki/Pareto_efficiency.
Wikipedia Tradespace [Online] // Wikipedia. - Wikimedia Foundation, Inc., May 2009j. - April 2010.

Wikipedia Vehicle Identification Number [Online] // Wikipedia. - Wikimedia Foundation, Inc., December 2009i. - December 2009. - http://en.wikipedia.org/wiki/Vehicle_Identification_Number.
World Health Organization International Classification of Diseases (ICD) [Online] // World Health Organization. - WHO, 2010. - March 2010. - http://www.who.int/classifications/icd/en/.

Acknowledgements

We thank Deb Bodeau, Martha Lorber and Leo Orbst for comments on earlier versions of this paper.