

# **PhD Comprehensive Exam**

Terrell Russell

University of North Carolina at Chapel Hill

School of Information and Library Science

Monday, January 25, 2010

## **Question 1 – Barreau**

Tagging has proved to be a useful tool for retrieving information. You recognize the value of tagging and the benefits of accessing the tags of others, regardless of their expertise, for finding desired information. Some have suggested that in digital environments traditional classification has little value since tags are more useful for retrieval. For knowledge management to work, organizations must be able to harness the capabilities within to create new knowledge, new products, and new services. What purposes do traditional classification systems serve in managing knowledge? Identify and describe at least two situations where traditional classification systems are beneficial, even in digital environments. Explain how making use of tags may affect the effectiveness of these systems.

# Response

## 1 Introduction

Organizations are created to formalize relationships. Relationships within the organization are recast as member or employee relationships. Relationships external to the organization are formalized between the organization itself and other organizations and people. This formalization of ties allows the organization to take on a role of stabilization, a role of record. The aggregation of voices that make up the membership of a new organization gain power in number and power from its newfound unity. An organization exists to give these voices a more powerful say in the world in which they have decided to band together.

Organizations then have a role to play when representing the voices and opinions of their memberships. The work of the organization is to further the goals and collective aspirations of its membership. In order to do this well, the organization must institutionalize, formalize, and organize the actions and activity of its membership. Determining a useful nomenclature and manner of classification of these actions and activity is a pivotal piece of succeeding as an organization.

In section 2, I will discuss the purposes of traditional classification systems with regards to an organization's motivation for managing knowledge. In section 3, I will discuss three examples where this type of classification is beneficial and useful. In section 4, I will talk about how our movement into a digital realm, with the possible addition of tagging, would affect the usefulness and effectiveness of these three systems.

## 2 Purposes

The purposes traditional classification systems serve in managing knowledge within an organization are many. Most of the declarative power of a classification system comes from the notion that it is speaking *for* the organization and has the implicit blessing of the organi-

zation as a whole. Because it has this institutional backing, the system carries more weight than if it was a system implemented by one person (even if it was, in fact, originally created by one person).

Because the classification system in question holds the authority of the organization itself, and because it is probably curated in some sense by experts who have spent time thinking about how it should be, the system **conveys some truth**. The order that is conveyed by the system is the order that has been blessed by the organization – it speaks on behalf of the organization. It therefore communicates prior work and prior knowledge. It saves work for the next generation of member or user as it has stored the work from before. The system is *efficient* in this manner.

A classification system also conveys a sense of **scope**. It communicates, *in toto*, a universe of what is known (or perhaps more specifically, what is important, what is cared about) by the organization. This scoping allows observers (internal and external) to better understand their work with regards to the precision and recall of documents and work that has come before them. They can know better how their work fits in with what has come before.

Additionally, a classification system, by definition, **defines a vocabulary**. When an organization has a working definition of what it means by certain words, how those words are represented in relation to other words, there is a sense of clarity conveyed to the reader. Defining accepted terms of art and having a working lexicon is very important to precision of communication both within and outside the walls of an organization.

One last purpose of a classification system is one of **social cohesion**. When the membership of an organization has confidence in what it knows, it spends less time running down the wrong paths. Having a shared scope and vocabulary give a unity of purpose to an organization. It gives them a sense of identity and helps differentiate themselves from others.

Organizations can have some of these traits without a formal classification system, to be sure. But when formalized, it is very clear that the mores and direction of a group of people

can be seen in their word choice and their structure.

Wang (2008) explains that Knowledge Management (KM) is constituted by four things: knowledge receptivity, knowledge absorption, knowledge sharing, and organizational memory. Classification plays a role in all four of these, but I feel it sits squarely among the last two. It both serves the role as sharing device and as memory store. Choo says that a Knowing Organization uses information in sensemaking, knowledge creation, and decision-making. A classification system provides structured information that can be used in all three of these activities. New information needs to be situated amongst the existing knowledge, juxtaposition with existing information may create new knowledge, and of course, moving forward, any decisions should be made in the context of what we already know (what is already stored in the classification system).

A classification system is mostly a store of externalized, explicit information. It cannot store what is not written or communicated. In this sense, classification systems are a formalized instance of what Nonaka called Combination in his SECI model (Socialization, Externalization, Combination, and Internalization). The storage, in plain sight, of what has already been codified is what classification systems are all about. To serve their roles well, good classification systems do not leave any tacit information unstated, they work exclusively with the explicit.

### **3 Examples**

In the following three examples labeled The Elements, Kingdoms of Life, and Library of Congress Subject Headings, I will quickly identify and explain how these systems work and how they are beneficial. I will then discuss how using tagging, or social labeling, could affect their effectiveness.

### **3.1 The Elements**

Our understanding of the chemical elements that make up our world is fairly deep. By definition, we organize the elements based on the number of protons in their nuclei. This is an integer count and we have successfully identified and studied all the elements from naturally occurring and ubiquitous Hydrogen (with one proton) to the far less stable, man-made super-heavy elements with over 100 protons. The simplicity of the definition of what constitutes a new and different element (proton count) allows wide agreement among scientists about the nomenclature and classification of elements. There are other facets as well. The weight of the atom (including neutrons), the melting and boiling points of the element, and the reactivity with other elements all constitute different ways to classify and organize the different elements. These are well known and well studied. There is little contention among the experts about how these things should be named and organized.

Because of this lack of contention, the system we have put into place serves us very well as we continue to learn and teach about chemistry to the next generation. As a classification system, the Periodic Table, based on the number of protons within the atom, is very solid. This affords great power in conveying a sense of order and understanding when faced with any new element or reaction. Since we know what has come before, we know how to expect a new situation to unfold.

The representation of this information is equally as useful in a digital context as it is in a physical context. What is really key here is the lack of contention.

### **3.2 Kingdoms of Life**

A second example of how a classification system is useful comes from a science higher up the chain. Above chemistry lies biology. As a science, biology also has many rules and understandings of how the world works. There are hundreds of years of experiments that show we mostly understand the things we have seen. However, as we continue to research the very big and the very small, we uncover new knowledge.

The system we have put into place to understand life itself has undergone some changes in the past few years. For a long time, this modern scientific system was called the Plant and Animal Kingdoms. All living creatures were part of one or the other. More recently, science (scientists) have begun to agree that there are actually five, no six, discrete kingdoms into which life can be classified. We now have plants, animals, fungi, and three kingdoms for small multi- or single-celled organisms. They are classified by such features as multi-cellularity, methods of movement, vertebrae, energy production, and reproduction.

As classification is supposed to be predictive, in that group members have similar characteristics to one another, the Kingdoms of Life are useful where we uncover new organisms that behave in new ways. We struggle with where to put a new one, and over time decide that it should be called a [something]. Of course, with new knowledge about how these organisms behave and reproduce, new kingdoms were slowly defined over time. Old knowledge was questioned and reformulated and repackaged to match our new knowledge of how the world worked.

A representation of an organism's kingdom/phylum/etc classification is useful in a traditional sense, but as just described, when recontextualization occurs, having the information in a digital format is more malleable and flexible. How many textbook chapters had to be reissued when Fungi was declared a Kingdom?

### **3.3 Library of Congress Subject Headings**

Moving up the spectrum of sociality and contention we arrive at a third example. The Library of Congress Subject Headings (LCSH) represent a vast array of distilled and disseminated knowledge from many sources and many years. The subject headings are used in nearly all US research libraries to label and classify the works held within. The general purpose is to communicate a standard language of topics so that related works may be more easily catalogued and discovered.

There are two places to see how this system plays out. First, at the classification system

level. Second, where the classification system is applied.

At the classification system level, the LCSH has grown over the years. It supposedly covers all of human knowledge, but of course, only so deeply. The medical and legal communities have developed their own subject headings as the LCSH does not delve deeply enough into those areas. This is a constant negotiation between breadth and depth of coverage for a system that is supposed to be universal. Of course, its stated mission is serving that of the United States and so is therefore not so universal from the start. Necessarily, there is a tension between changing quickly enough to facilitate new language and new realities and being stable enough to not “break” existing classifications. The LCSH do change, but it is a slow and very human-driven process.

More specifically, as a success, we could look at the agreement level for certain works across different libraries. As each library catalogues its holdings individually, there is a natural “score” of how similar each library classified its works. Where works are catalogued similarly, we could claim that the system is doing a great job and successfully capturing the essence of the work being described. Some other works may not be catalogued similarly across the board. This may mean that the work is less understood, or that the subject matter itself is more contentious.

Regardless, the cataloguing of the work is a statement by the organizing body (in this case, a library). It serves a purpose of organizing what is known as well as placing a work within a known universe of understanding.

## 4 Tagging

As the LCSH are more social in nature than either Kingdoms of Life or the count of protons in each element, there is more contention about what these subject headings should be and how they should be defined as related to one another. Of course, experts are making these decisions, but even so, these experts do not always agree. I spoke about probabilistic facts

in my literature review and pointed out that facts are defined as where the experts no longer disagree. Through observation and discussion, experiment and analysis, they have agreed that something is safe enough to be bottled and put on the shelf as “true”. If some new information comes along that makes them question that truth, then they may take it back down and work on it some more until the contention is gone again.

Traditional classification systems do not handle this recontextualization very well. They are slow and methodical (by design) and therefore are good at capturing what is known and what is agreed.

Tagging adds a flexibility and possibility to the classification of information where it is not as agreed upon. Where there is contention, we can see it in the tags. When multiple experts are allowed their own voices, we can clearly see where the contention lies, where there is overlap, and where there is not.

If we were to add tags to the Periodic Table, we would probably not learn very much. There is little contention there and little to argue about. The vast majority of discussion in physics is not at the elemental level anymore. That knowledge has been bottled and put on the shelf. It is known.

Adding something like tagging to the Kingdoms of Life would be interesting - although, frankly, that is basically what we have now, but in a very distributed, slow manner – called “all of biological research”. Distilling that work down to simple tags would be a mammoth job, and probably provide us, still, with little tangible results. Biology is still far down the empirical side of the social spectrum.

Adding tagging to the LCSH process is more interesting. As it is the most social of the three discussed, it plays most strongly into the hands of what tagging is supposed to be good at. I suspect that if the Library of Congress were to create a project and allow subject matter experts from around the world to label and freely associate the subject headings within the current LCSH, we would see a lot of overlap with the current system. But I suspect we would also see, very clearly, where there is large agreement around a certain few terms that

are either outdated or misplaced in the broader term (BT) or narrower term (NT) hierarchy. No doubt the eyes of many would make these types of “bugs” visible. And arguably, where the consensus has moved, the LCSH should follow suit.

This would be a fascinating case study, and a project that I would love to see happen. I suspect if it was done well, a proper hybrid approach including what is good about the existing LCSH and what is provided by tagging, we would have a much better LCSH than is currently in place. The system would have less contention, perhaps, and the libraries (users) of the system would have greater confidence in the system as a whole.