

The Strange Case of Eugenics: A Subject's Ontogeny in a Long-Lived Classification Scheme and the Question of Collocative Integrity

Joseph T. Tennis

The Information School, Box 352840, Mary Gates Hall, Suite 370, University of Washington, Seattle, WA 98195-2840. E-mail: jtennis@uw.edu

This article introduces the problem of collocative integrity present in long-lived classification schemes that undergo several changes. A case study of the subject "eugenics" in the Dewey Decimal Classification is presented to illustrate this phenomenon. Eugenics is strange because of the kinds of changes it undergoes. The article closes with a discussion of subject ontogeny as the name for this phenomenon and describes implications for information searching and browsing.

Introduction

Classification schemes are built at a particular point in time; at inception, they reflect a worldview indicative of that time. This is their strength, but results in potential weaknesses as worldviews change. For example, if a scheme of mathematics is not updated even though the state of the art has changed, then it is not a very useful scheme to users for the purposes of information retrieval. However, change in schemes is a good thing. Changing allows designers of schemes to update their model and serves as a responsible mediator between resources and users. But change does come at a cost. In the print world, we revise universal classification schemes—sometimes in drastic ways—and this means that over time, the power of a classification scheme to collocate is compromised if we do not account for scheme change in the organization of affected physical resources. If we understand this phenomenon in the print world, we can design ameliorations for the digital world.

There is much discussion in the research literature about how schemes could be changed to benefit the user or to better represent a particular view on the domain. In some cases, researchers approach this problem by documenting

cases where schemes break down in relation to current conceptions about identity. For example, Bowker and Star (1996, 2000) documented how difficult it is to create a scheme for nursing work because nurses work at the intersection of both medical and caregiving professional knowledge. They noted that these extant classification schemes are very difficult to change because of inertia (Bowker & Star, 1996, 2000). Olson (2007) discussed how current systems cause scatter because of biases in the structure and semantics of the scheme. This scatter is different than that caused by scheme change. Rather, it is built into any single instance and is tied to interpretations of literary or user warrant of the domain, not about how semantics change and the effect of that on collocation. With regard to the subject of lesbians, she stated that the revised edition of the Dewey Decimal Classification eliminated two places for lesbians. In the previous version, lesbians could be considered as women or as gay. In the revised edition, they were collapsed into one class, negating the possibility, according to Olson, that lesbians could be both a subclass of homosexual and a subclass of women (pp. 535–536).

Here, subject scatter (lesbians) has changed to a single collocation based on a reinterpretation of literary warrant. In Olson's (2007) article, she documented the undesirable logic of conflation. We are not, in our work here, judging the merits of expanding or collapsing classes, though we have documented their types and varieties of such structural (and concomitant semantic) changes elsewhere (Tennis, 2007). Olson's example provides us with what we have called a structural change (Tennis, 2007, pp. 90–91). A structural change is one of many types, and in this example, we only see a single change. The questions posed here are "What is the shape of this change over a longer period of time, and what are the contours of change in a long-lived scheme?"

In the following sections, we will describe the purpose of classification and how it is challenged with changes made to the scheme. Next, we describe a case study of scheme

Received October 31, 2011; revised February 26, 2012; accepted February 27, 2012

© 2012 ASIS&T • Published online 23 May 2012 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.22686

change, and then use this case study to illustrate the concept of subject ontology (i.e., the life of a subject over time). We close with a discussion of the implications of designing classification schemes with subject ontology in mind.

Classification for Purposes of Organization and Information Retrieval

The purpose of bibliographic classification is to collocate kinds of documents based on an interpretation of their subjects. It does this by creating a system of classes, arranged by kind, generally into hierarchical and systematically ordered arrays. A major concern of 20th-century research on classification has been the preservation of a scheme's integrity, even with the addition of new classes. That is, theorists have concerned themselves with the evolution of a scheme in one particular sense: They want to maintain collocative integrity of a scheme while adding to it. This has led to the development of faceted and analytico-synthetic classification (cf. Aitchison, Gilchrist, & Bawden, 2000; Bliss, 1939; La Barre, 2000; Ranganathan, 1967; Vickery, 1960, 1966).

However, much of this work has been focused on structure. Ranganathan (1967) was concerned with the structure of a classification that is robust enough to handle new subjects and place them into the appropriate hierarchical and systematic relationship with extant classes. According to his design requirements, the collocative integrity of the scheme should be maintained even when new classes are added. To add new subjects, one needs to break apart the subject and recombine it according to principles that apply to every facet of the subject. This ensures that when new subjects (made up of facets) surface, we can position them in a sequence helpful to the user, not pasted on the end after the Miscellaneous Class.

Subject Across Time: Ontogeny

While many theorists have concerned themselves with how to design a scheme that can handle the addition of subjects, very little has been done to study how a subject changes after it is introduced to a scheme. Simply because we add civil engineering to a scheme of classification in 1920 does not signify that it means the same thing today. Almost 100 years have passed, and many things have changed in that subject. We may have subdivided this class in 1950, thereby separating the pre-1950 meaning from the post-1950 meaning and also affecting the collocative power of the class civil engineering. Other classes in the superclass of engineering might be considered too close, and are eliminated over time, affecting the way the classifier does her or his work (cf. Tennis, 2007; Tennis & Sutton, 2008). It is because of these concerns, coupled with the design requirement of collocation in classification, that we need to look at the life of a subject over time—the subject's scheme history or ontogeny.

To explore this notion of an ontogeny in the context of scheme change, we will examine the subject of eugenics in the Dewey Decimal Classification (DDC), from its first

appearance in the seventh edition in 1911 to the present 23rd edition. Understanding the ontogeny of a subject is instrumental to maintaining the design requirements of any classification scheme as it evolves.

However, as we will show, not everything remains collocated when a long-lived scheme such as the DDC changes over time and the positioning of associated physical resources in a particular collection fail to keep up with those changes. As editors rework the scheme, classes get moved and are then assigned new numbers. Even with written policies in place, the inevitable failure of collocation sometimes happens. Over time, the scheme uses the same number for two different subjects, thereby rendering the design requirement of collocation confounded, if not useless. In the case of eugenics, there are collections that collocate books on eugenics studied from a biological perspective with those on the reproductive parts of plants because both have the same number, 575.6.

If the *raison d'être* of classification schemes in knowledge organization systems is collocation of kinds of documents based on an interpretation of their subjects, we must understand how scheme change affects subjects in schemes over time. This article reports one facet of descriptive work in this area to improve design and maintenance of classification schemes as changes occur over time.

Approach and Goals

The goals of this article are to (a) describe a particular instance of the structural and semantic changes to DDC; (b) document a subject's ontogeny in the process; and (c) point to potential implications of these changes on retrieval, specifically the specification that a classification scheme needs to maintain its collocative integrity over time. The argument advanced here is that a better understanding of change will allow designers of systems to accommodate change, thereby maintaining—over the long term—the functional requirements established by the structure and semantics of the scheme. The subject of eugenics is a valuable case study because there are a number of types of changes made over the course of this subject's ontogeny. First, the worldview around eugenics changed from 1911 to the present. This is reflected in the publication patterns (numbers and type) and sentiment about the validity of eugenics as a scientific pursuit. This change in sentiment can be seen in the recasting of eugenics in different disciplines in the DDC and the rhetoric of publications across those dates. Finally, were we to take a common (rather than strange) case, say anatomy, we would not see very much change and therefore not see the challenges to collocation and collocative integrity of a scheme through time (Tennis, Thornton, & Filer, 2012). Anatomy is the opposite of strange in the DDC, and though it has an ontogeny, it is not instructive for the study of the challenges to collocative integrity.

Second, the DDC accommodated this change in literary warrant in a diverse set of ways: lumping and splitting classes and reorganizing the topic along different

disciplinary lines. We can observe each of these through use of the tables extracted from the DDC schedules.

The purpose of this article is not to demonstrate how collocative integrity has affected cataloger decisions. This work is a separate endeavor, with reports on this data surfacing in Tennis et al. (2012). Because of the complex nature of this topic, we present one conception here, collocative integrity. The concept of collocative integrity is perhaps more directed to the designer's and editor's views, not the cataloger's view.

To achieve the goals set out in this article, we approach the investigation of subject ontogeny and the construction of the concept of collocative integrity from a neo-pragmatic framework, which claims that words are tools for action (Rorty, 1982, 1999). In this particular case, we also are informed by Berger and Luckman's (1967) work on social constructionism, and the tenet that once something is constructed, say a classification scheme, it can be used in many different ways and thus must be understood independent of its creators. That is, we hold that a subject has a life of its own (metaphorically), and we need to better understand that life to critique the value and functionality of long-lived classification schemes that change over time.

An analogy might be that we understand aerobic or anaerobic mechanisms of the body so that we can be better at helping others with their fitness routines. A basic knowledge of the underlying mechanisms of a system, whether it is blood/oxygen exchange or subject ontogeny, will allow us to make informed changes to the routines that relate to that system. For example, we can change workout regimes, or in the case of classification, change our work practice of classing and changing classification schemes.

This approach is useful for us because it (a) acknowledges the language at work in classification and (b) provides for a way in which we can handle the object of study as separate from its creators and separable from its history.

This work is a case study, one of hopefully many more that can document the varieties of change in schemes, operating at a level of detail that makes clear our assumptions about semantics and structure over time. We believe this is an advancement of our understanding of schemes and the ramifications of change based on an assumption about the *raison d'être* of classification schemes. There is a neo-pragmatic functionalism in this approach; namely, that the use of a particular kind of language (the DDC) to collocate documents in a collection over time is a pragmatic act. In this perspective, the DDC is an assemblage of words used for action. Because it is a pragmatic act, the action of collocation is the beginning and end of our concern. How does the designer of a scheme collocate kinds of subjects? If the subject is in one place for its entire life in the scheme, there is no question about its collocative integrity. If it changes, and changes dramatically, then we have to signal this to the designers and editors in a particular way, such that the functional requirements of classification are retrained. Thus, we need to study the nature of these changes case by case, assemble a body of knowledge around these cases, abstract

from them a theory (or set of theories) about how change occurs, and from here, design interventions that could help designers of schemes.

Finally, following Rorty (1982, 1999), we posit that the findings from a neo-pragmatic study of subject ontogeny will be a set of vocabulary useful for understanding and designing for scheme change.

DDC: The Case of Eugenics

The DDC has gone through 23 editions in its full form. First authored by Mevil Dewey in 1876 as a pamphlet that listed a few classes of subjects, the DDC has grown into a multivolume work. Many editors have been involved in the evolution of the scheme (Comaromi, 1976), and coupled with the evolving nature of subjects published, it has changed over time—sometimes dramatically.

Miksa (1998) described three major periods of change: (a) Beginnings, which spans Editions 1 to 6 (1876–1899); (b) Conflict, spanning Editions 7 to 15 (1911–1953); and (c) Recovery and Advance, which contains Editions 16 to 21 (1958–1996) (p. 5). Our case study moves from Conflict to the Recovery and Advance periods, 1911 to the present.

The DDC is primarily an enumerative scheme organized by discipline, meaning that most of the classes are explicitly listed. The classifier does not need to create them through synthesizing numbers (though it is possible to do so), and topics are organized by broad disciplinary categories such as Religion, Social Science, Science, and Literature. This means that a topic could appear in more than one place if it is studied from more than one disciplinary perspective. In many cases, topic regularly lines up with discipline: For example, anatomy is almost always studied in Biology and Art. And of course, the researcher interested in ontogeny also has to account for the philosophy of . . . or history of . . . questions of topics which would place them in Philosophy or History regardless of whether they are horseshoes or hand grenades.

However, not all topics make sense in this construction of topics and disciplines. For example, at its first appearance in the DDC, eugenics is a biological science. This is not a bizarre placement in 1911, but is so now (to many audiences), just as anatomy, which can be studied from any discipline, is so closely linked with biology and art that it is most commonly discussed in biology books and art books, and most commonly classed in Biology or Art. The case of eugenics is strange, for the reasons mentioned earlier but also because it has changed, within the scheme, from a science to a nonscience. The reorganization is a redefinition in a forceful way, such that even if there were a book on eugenics written from the biological perspective, it could not go into the life sciences in the same position—thereby negating outright any collocative integrity. The DDC reflects this change quite clearly because of its disciplinary organizational structure. So although it was once possible to say through the lens of the classification scheme that there are

books published on eugenics as a science (specifically life science), it is no longer possible to do so.

This particular case of change is strange, too, because of the amount of change in the ontology of eugenics. As Miksa (1998) noted:

The pursuit of classificatory change of the kind involving relocation, when viewed on an edition-by-edition basis, has actually been pursued with a great deal of constraint. Constraint has been exercised in part because of the potential disruption relocations have for the libraries which use the system, especially for shelf arrangement. (p. 25)

He goes on to emphasize the rarity of “wandering subjects”:

Constraint has been exercised because some changes need more compelling evidence before being adopted than might at first seem obvious and this has the effect of delaying decisions about them. This is done to guard against hasty decision about changes that might eventually have to be remade. In the worst cases, decisions of the latter kind have led to the specter of “wandering subjects.” Wandering subjects are subjects which have moved from one place to another, and sometimes back and forth between the same two places, over more than one edition of the system. (pp. 25–26)

Eugenics displays this “wandering” nature and other traits that make it many degrees “stranger” than is anatomy in its life and in the lives of catalogers from 1911 to the present.

Scheme Change

In general, we understand what happens to classes when classification schemes change (Miksa, 1998, pp. 23–27). Editors change the structure of the scheme, change the words used to navigate and understand the structure, and respond to a change in literary warrant by eliminating or redefining a class (Tennis, 2007). The ramifications of change can be small or great, and there are a number of ways that we might study these ramifications. One way is to carry out an in-depth case study to see how a topic is handled over time. In the next section, we chart how eugenics is handled in the DDC. In so doing, we establish the ontology of the subject.

What we find in this case is significant structural change over time as eugenics moves from being a Biological Science to something of a policy of thought or an ideology akin to racism. We see the class numbers in the 500s (Sciences) disappear at some point post-WWII. If discussed at all post-WWII, such discussion of eugenics takes place in History or Social Sciences.

Eugenics as Subject

Eugenics has many definitions, and the acceptability of those definitions has changed over time. The term eugenics

comes from the Greek roots for “good” and “generation” or “origin” and was first used to refer to the “science” of heredity and good breeding in about 1883 (People and Discoveries, 1998). Galton (1883) noted that human eugenics is the investigation “of the conditions under which men of a high type are produced” (p. 44). For the purposes of the 1921 conference on eugenics, the discipline was defined as “the self direction of human evolution” (Laughlin, 1923, p. 21).

In general, eugenics is the study of bettering the human race through scientific investigation into selective breeding. In the popular imagination, it is currently linked with Nazi science and the pursuit of a perfect race. What is interesting about this subject is that it is no longer possible to consider it a Biological Science in the eyes of the DDC.

In the past, eugenics was part of biology. As such, there were many publications on the scientific improvement of the human race through breeding for desirable traits. Its first appearance in the seventh edition in 1911 placed it near the discipline of Genetics in Biology, and it remained in that class until the 1950s. This means that all books assigned a class number under eugenics from 1911 to the 1950s could be classed in the sciences. However, after that, if all libraries were using the most up-to-date edition of the scheme, they would have had to use another class number.

We now present three tables that outline the ontology of eugenics in DDC. The first table shows spaces available for eugenics organized in chronological order by decade. It lists the class numbers (where eugenics could be placed), edition dates (when the class numbers were issued), edition numbers, and decade in which the editions were issued.

In Table 1, we see that eugenics appears in the 1910 decade. Classifiers have three choices during that decade (575.1, 575.6, and 613.94), and there were four editions of the classification scheme (Editions 7–10).

From the seventh edition on, each decade provides its unique constellation of class numbers, with 1980 being the decade to offer the most classes. This is one approach to viewing the ontology of the subject eugenics in the DDC.

The next two tables chronologically view the ontology, but this time from the spatial metaphor afforded by the numbers and without the edition dates or edition numbers. Table 2 shows each individual number, and a darkened box shows when eugenics appears in that class. Table 3 shows the “centuries” (broad classes at the top of the hierarchy), and the darkened boxes show when eugenics appears in the class. This is a second approach to visualize the ontology of eugenics in the DDC.

As expected, we see interruptions, erasures, and movement in the subject’s ontology. If it were not a strange case, we would expect to see very little change. However, as expected, we see how this case causes challenges for searching on a particular class number or browsing a particular section of the collection. The interruptions happen in every century of the classification scheme (100s, 300s, 500s, and 600s). Those interruptions damage the collocative integrity of the extant class numbers. We see erasures as well; there are points when classes surface and then disappear. The

TABLE 1. "Eugenics" in the Dewey Decimal Classification Relative Index.

Decade	1870	1880	1890	1900	1910	1920	1930
Class Numbers	[none]	[none]	[none]	[none]	575.6, 575.1, 613.94	575.1, 613.94	136.3, 159.9223, 575.1, 613.94
Edition Dates	1876	1885, 1888	1891, 1894, 1899		1911, 1913, 1915, 1919	1922, 1927	1932
Edition No.	1st	2nd, 3rd	4th, 5th, 6th		7th, 8th, 9th, 10th	11th, 12th,	13th
Decade	1940	1950	1960	1970	1980	1990	2000
Class No.	136.3, 301.323, 364.3018, 364.42, 575.1, 613.94	301.323, 364.42, 612.63, 613.94	613.94*	174.25, 363.92, 363.98, 364.4, 364.42, 613.94	174.25, 323.97, 362.36, 363.92, 363.97, 363.98, 364.42, 613.94	174.25, 362.36, 363.92, 363.97, 364.4, 613.94	176, 323, 362.36, 363.92, 363.97, 364.4, 613.94
Edition Dates	1942	1951, 1958	1965	1971, 1979	1989	1991	2003
Edition No.	14th	15th 16th	17th	18th, 19th	20th	21st	22nd

*Eugenics is not even in the Relative Index, although it is possible to still use this class for the meaning of eugenics.

striking example is in the Biological Sciences where four decades of class numbers completely disappear in subsequent editions; this happens in other centuries of the scheme as well.

We also see the challenges for interpreting the data. In some editions, the relative index does not have an entry for eugenics, and this requires us to examine the class number and judge whether the topic eugenics is still preserved. When we examine catalogers' decision making in regard to this change, we would be sensitive to what is already in their collection. This is another frontier for future research.

A View on Literary Warrant in Relation to Classes in the DDC

An important consideration for subject ontology is its relationship to literary warrant. How much was published with words (and hence an indication of the topics) such as eugenics? We can look at the words in the schedule and relative index related to each class, and see how it appears and disappears in books over time by searching in Google Books through the Ngram Viewer (Michel et al., 2011), which contains 5.2 million books in seven languages, or approximately 4% of all books ever written (Michel et al., 2012). The Ngram Viewer simply looks for the appearance of words in the text of digitized books held by Google. The purpose of the tool is to trace the number of occurrences of a term (more specifically, a string of characters and spaces where necessary) in books over time.

In the visualization in Figure 1 using the Google Ngram Viewer, we see the frequency of the word "eugenics" in the corpus of digitized books in Google's database. The y axis shows us the percentage compared to the whole corpus; the x axis shows the dates between 1909 and 2008. Thus, we can see how often the term eugenics appears in this corpus.

It is instructive to see the term "eugenics" fall out of favor by the 1960s and then rise in frequency after the 1990s. We also can look at the phrase "human eugenics" in the Ngram Viewer (see Figure 2).

Here, we see that the phrase is more popular before the 1950s, at which time it drops out of discussions, only to rise a little in the 1960s. We gain some modicum of context when we compare these data to the tables of DDC numbers.

The Place of Eugenics in the DDC

Langridge (1989) distinguished between discipline and topic in the context of the DDC using horses as his example. As a topic, horses can be studied from multidisciplinary points of view, meaning for a searcher that horses will be found in multiple places in the DDC. This is theoretically possible for any topic, but is less likely to happen with some of them. For example, anatomy might be studied from multidisciplinary viewpoints, but in practice, is usually fixed to the Applied Sciences and Art disciplines in the DDC (Thornton & Tennis, 2010).

TABLE 2. "Eugenics" in the Dewey Decimal Classification Classes organized by decade (same data as in Table 1).

		1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
Useful Arts (later Technology)	613.94	[Shaded]									
	612.63	[Shaded]									
Sciences	575.6	[Shaded]									
	575.1	[Shaded]									
Social Sciences	364.42	[Shaded]									
	364.4	[Shaded]									
	364.3018	[Shaded]									
	363.98	[Shaded]									
	363.97	[Shaded]									
	363.92	[Shaded]									
	362.36	[Shaded]									
	323.97	[Shaded]									
	323	[Shaded]									
	301.323	[Shaded]									
Philosophy	176	[Shaded]									
	174.25	[Shaded]									
	159.9223	[Shaded]									
	136.3	[Shaded]									

The 17th edition from 1965, the only edition issued in the 1960s, did not include a Relative Index. However, it was still possible to class books in 613.94 and still carry the meaning of eugenics. Therefore we note the difference with a lighter shading, rather than no shade.

TABLE 3. "Eugenics" in the Dewey Decimal Classification centuries organized by decade (same data as in Tables 2 and 3).

		1910	1920	1930	1940	1950	1960	1970	1980	1990	2000
History (later History and Geography)	900s	[Shaded]									
Literature	800s	[Shaded]									
Fine Arts (later Arts and Recreation)	700s	[Shaded]									
Useful Arts	600s	[Shaded]									
Sciences	500s	[Shaded]									
Language	400s	[Shaded]									
Social Sciences	300s	[Shaded]									
Religion	200s	[Shaded]									
Philosophy (later Philosophy and Psychology)	100s	[Shaded]									
Generalities	000s	[Shaded]									

The 17th edition from 1965, the only edition issued in the 1960s, did not include a Relative Index. However, it was still possible to class books in 613.94 and still carry the meaning of eugenics. Therefore we note the difference with a lighter shading, rather than no shade.

With eugenics, we see that it is a topic that can be expressed using different disciplinary languages and therefore is classed in different places in the DDC. However, we see with eugenics a discontinuation of classes—a rendering of certain disciplinary voices silenced in the scheme as it goes through various editions.

This is similar to the phenomena observed by Bowker and Star (2000), Olson (2007), and Furner (2007). In each of these studies, we see how classification schemes ossify decisions based on bias and power. Olson's critique reinforces our assumptions about the power of collocation and the need to understand collocative integrity of a scheme over time. For Olson, collocation must be in accordance with principles of equity. For example, she described the fact that mental illness in Library of Congress Subject Headings is considered male and requires reclassification of concepts for mentally ill women, children, and older people. The silent assumption is that men are the default universal, and its only

through naming women, children, and older people that we see the power and need to represent subjects in the scheme.

What Is a Subject's Ontogeny?

From this investigation we might draw a few conclusions about what a subject's ontogeny is and why it matters. First, indexing languages like classification schemes are interventions, designed with purpose. That is, people build them to do some sort of work. The broadest possible purpose for these interventions is to help people find information. This is fine, as far as it goes, but it does not go very far. Classification schemes help users by listing subjects in classes that are hierarchical, mutually exclusive, jointly exhaustive, and systematically ordered. Generally the hierarchies start, at the top level, with disciplines such as science, history, and so on. The systematic order of each rank is meaningful in some way. Ranganathan (1967) cited some 30 approaches for

Graph these case-sensitive comma-separated phrases: between and from the corpus with smoothing of .

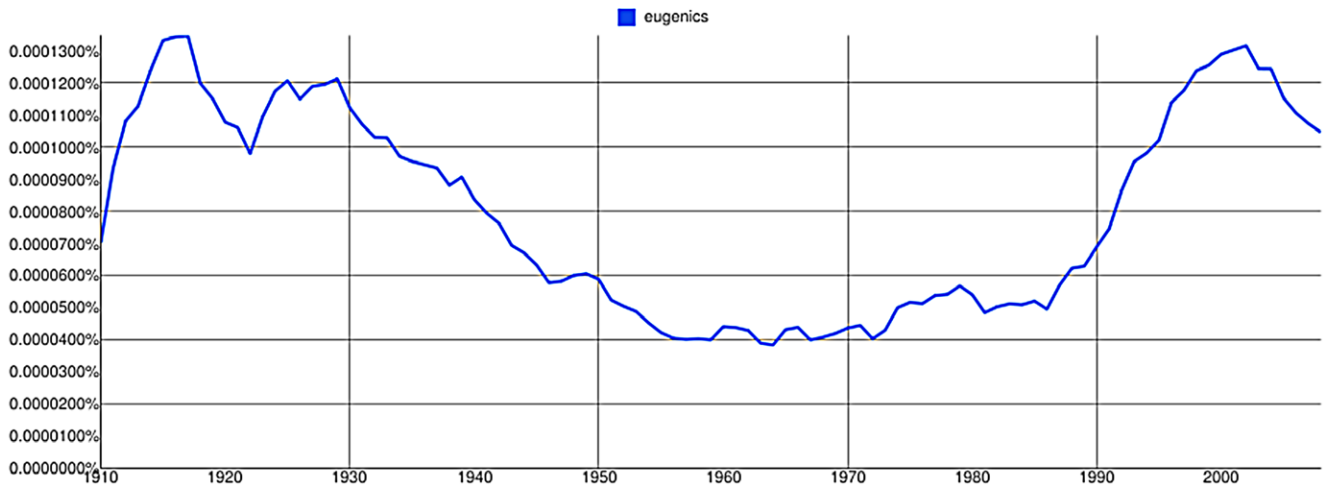


FIG. 1. Google's Ngram display for "eugenics" 1910–2008. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Graph these case-sensitive comma-separated phrases: between and from the corpus with smoothing of .

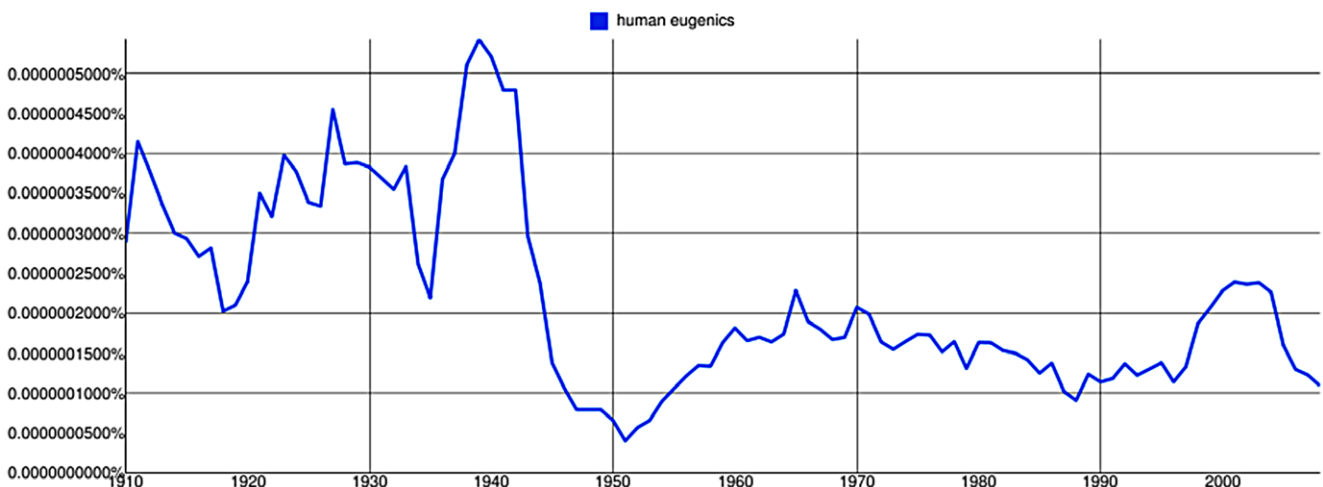


FIG. 2. Google's Ngram display for "human eugenics" 1910–2008. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

example, a ranked array might be arranged from oldest to youngest, from smallest to largest, or some other meaningful order. Finally, in its ideal form, a classification scheme should (a) have one place for every subject, making it mutually exclusive, and (b) have a place for every subject represented in the collection, making it jointly exhaustive. These design requirements function to disambiguate subjects represented and save time for the reader by providing one place to look if the subject is represented at all in the collection.

Leaving aside the problem of assigning a document to one and only one class (cf. Mai, 2001) and all the proposed ameliorations (e.g., Hjørland, 1992; Mai, 2005), the purpose of the scheme remains. What follows from this ideal conception of subject classification is that once we have a place for a subject—say 575.1—it should be preserved over time. As we demonstrated earlier, that is not the case for long-lived schemes such as the DDC. Thus, we immediately want to know how this affects the ability of the scheme to fulfill its purpose as an intervention.

If there were no change from 1911 to 2011, then the ontogeny of the subject eugenics in the DDC would be linear—a straight line from start to finish. Explicitly, it means that 575.1 would always be the address for eugenics. This means that we would expect all of the resources classed from 1911 to 2011 to be collocated, no matter when they were classed.

Since there is change, the subject ontogeny has many characteristics: (a) It branches, moving from one class to two classes; (b) it steps, in that it moves from one class to another within the same superclass; (c) it converges, in that it either eliminates or combines classes over time; and (d) in one decade, it becomes a shadow, where the term disappears but the class exists. Since the scheme changes and we assume that classifiers are not reclassing resources in the current class, this means that the searcher is (a) relying on terminological mechanisms and not the class structure of the classification scheme to access the documents, and (b) the searcher will find (potentially) one class with different subjects (e.g., 575.6, which is now the two different subjects depending on which edition of the DDC is consulted). It is either eugenics as studied in biology or the reproductive parts of plants as studied in biology.

The subject's ontogeny is the life of the subject in the scheme. But the ontogeny of the subject in the scheme has implications for the resources classed and yet to be classed. Because the designers of classification schemes want to revise to represent the most current understanding of the domain, the classifier then has two sets of resources in their collection: those classed before the current edition and those classed after the current edition. It is an empirical question as to whether the more a scheme changes, the more likely the classifier is to change to classing resources using the most current location afforded by the scheme. However, it is clear that in a long-standing collection, the collocative integrity of the underlying scheme is lost when the scheme changes and there is no reclassification of resources activity.

From a designer's point of view, you do not want a class to be forever in one place. It may be and, in fact, has been the practice to destroy large sections of the DDC to more faithfully represent the domain (e.g., see Music; DDC, 20th edition). These destruction/resurrections are called phoenixed schedules, and ideally represent a more cogent and up-to-date conception of the domain (Olson, Boll, & Aluri, 2001). Nevertheless, scheme change—whether subtle or drastic—is a design constraint that we must account for because it affects how the user interacts with the resources over time.

Implications

The Need to Consider Change

While much excitement about indexing languages and categorization schemes exists around new and emerging information organization and representation systems such as

ontologies, social tagging sites, and Wikipedia's category system, we can learn much from a closer examination of older systems. This is especially true if we want to better understand the effect of change on these kinds of systems over time. Systems have history. They appear at one point in time, and then worldviews march on. Designers of schemes must account not only for the scheme at one point in time but also for the diachronous (through time) nature of schemes. Information access is not only a synchronic concern but also a diachronic concern. That is, we catalog or index at one point in time predicting future use of the cataloging or indexing system. If the reason why a system is built in 1876 is not maintained to the current day, we cannot expect to evaluate it using the same constructs.

Designing for Change

A second consideration is designing for change. By investigating a subject's ontogeny, we more clearly understand what kind of metadata is required in our records for particular classes in our schemes. What information about classes do we need to accommodate revisions to our scheme? Work has begun in this area (Tennis, 2007; Tennis & Sutton, 2008), specifically in the building out of different aspects of term records for managing a thesaurus and thesaurus-like indexing languages. Likewise, systems designers have begun to consider these design specifications to track changes in their knowledge organization systems (e.g., Eito-Brun, 2011; Hillmann, Coyle, Phipps, & Dunsire, 2010; Panzer & Zeng, 2009; Tuominen, Laurene, & Hyvönen, 2011). However, more case studies are required to identify the similarities and differences between different indexing languages and classification schemes.

Finally, the retention of conceptual meaning through time is of import to many digital information systems that have long-term preservation as their mission (Tennis, 2002). Studying a subject's ontogeny will help those systems maintain their value over time.

Future Work

There are open questions about degree, rate, and impact of change on long-lived schemes. What effect does this have on the functionality of classification? What does this do to the concept of sharing data in linked data environments predicated on other structures such as the Machine Readable Cataloging record?

Also, are different subject ontogenies typical? That is, is eugenics different from horses, or anatomy, and the like? Comparative work will help us better understand the nature of subject ontogeny itself and its effect on collocative integrity.

Finally, work could proceed on the measurement of collocative integrity and benchmarking scheme change, specifically with regard to a subject's ontogeny. This would allow editors of schemes to make informed decisions about the

changes they make to schemes vis-à-vis the functional requirement of collocation through hierarchy and systematic arrangement.

Conclusions

The strange case of eugenics is instructive as a case study of subject ontogeny and is informative with regard to the problems of collocative integrity in a long-lived classification scheme. It is instructive because we have begun to document erasures and interruptions; we can now look for these in other cases. As noted by Furner (2007), who describes how race is handled in the DDC, our concerns for representing people in classification have changed dramatically in the last century. Yet, there is a legacy of literature, identified through our long-lived classification schemes. Buckland (2012) also alluded to this problem:

Meanings are established by usage, and so always draw on the past. The librarian, then, is creating descriptions by drawing on the past, but expressing them with an eye to the future. This Janus-like stance might seem difficult enough in a stable world, but reality is made much worse by time, by technology, by the nature of language, and by social change. (p. 156)

Understanding a subject's ontogeny leads us a long way toward understanding the power of collocation through representation, and the weight of meaning in the value added through classification and indexing. Identifying the contours of this problem helps us maintain a scheme's purpose—its collocative integrity.

Acknowledgments

I am grateful to Stuart Sutton, Andrew Filer, Katherine Thornton, Tod Robbins, Gary Gao, and Monica Carraway for their help and comments on this work. I am indebted to the reviewers, whose suggestions made this a significantly richer article.

References

Aitchison, J., Gilchrist, A., & Bawden, D. (2000). *Thesaurus construction and use: A practical manual*. London, UK: Fitzroy Dearborn.

Berger, P.L., & Luckman, T. (1967). *Social construction of reality: A treatise in the sociology of knowledge*. Garden City, NY: Anchor Books.

Bliss, H.E. (1939). *The organization of knowledge in libraries and the subject-approach to books*. New York, NY: Wilson.

Bowker, G., & Star, S.L. (1996). How things (actor-net)work: Classification, magic and the ubiquity of standards. *Philosophia*, 25(3-4), 195-220.

Bowker, G., & Star, S.L. (2000). *Sorting things out*. Cambridge, MA: MIT Press.

Buckland, M. (2012). Obsolescence in subject description. *Journal of Documentation*, 68(2), 154-161.

Comaromi, J.P. (1976). *The eighteen editions of the Dewey Decimal Classification*. Albany, NY: Forest Press.

Eito-Brun, R. (2011). Integration of SKOS and SRU in a distributed collaboration environment for archival material description. In A. Gilchrist

& J. Vernau (Eds.), *Facets of knowledge organization. Proceedings of the International Society for Knowledge Organization (ISKO) United Kingdom Biennial Conference* (pp. 375-384). Available at: <http://www.iskook.org/conf2011/papers/eitobrun.pdf>

Furner, J. (2007). Dewey deracialized: A critical race-theoretic perspective. *Knowledge Organization*, 34, 144-168.

Galton, F. (1883). *Inquiries into human faculty and its development*. London, England: Macmillan.

Hillmann, D., Coyle, K., Phipps, J., & Dunsire, G. (2010). RDA vocabularies: Process, outcome, and use. *D-Lib Magazine*, 16(1/2). Available at: <http://dlib.org/dlib/january10/hillmann/01hillmann.html>

Hjørland, B. (1992). The concept of "subject" in information science. *Journal of Documentation*, 48(2), 172-200

La Barre, K. (2000). Bliss and Ranganathan: Synthesis, synchronicity or sour grapes? In C. Beghtol, L.C. Howarth, & N.J. Williamson (Eds.), *Dynamism and stability in knowledge organization. In Proceedings of the Sixth International Conference of the International Society of Knowledge Organization, Toronto Canada* (pp. 157-163). Würzburg: Ergon Verlag.

Langridge, D.W. (1989). *Subject analysis: principles and procedures*. New York, NY: Bowker-Saur.

Laughlin, H.H. (1923). The Second International Exhibition of Eugenics held September 22 to October 22, 1921, in connection with the Second International Congress of Eugenics in the American Museum of Natural History. Baltimore, MD: William & Wilkins.

Mai, J.-E. (2001). Semiotics and indexing: An analysis of the subject indexing process. *Journal of Documentation*, 57, 591-622.

Mai, J.-E. (2005). Analysis in indexing: Document and domain centered approaches. *Information Processing & Management*, 41, 599-611.

Michel, J.-B., Shen, Y.K., Aiden, A.P., Veres, A., Gray, M.K., Pickett, J.P., . . . Aiden, E.L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014), 176-182.

Michel, J.-B., Shen, Y.K., Aiden, A.P., Veres, A., Gray, M.K., Pickett, J.P. et al. (2012). *Cultureconomics*. <http://www.cultureomics.org/>

Miksa, F. (1998). *The DDC, the universe of knowledge, and the post-modern library*. Albany, NY: Forest Press.

Olson, H.A. (2007). How we construct subjects: A feminist analysis. *Library Trends*, 56, 509-541.

Olson, H.A., Boll, J.J., & Aluri, R. (2001). *Subject analysis in online catalogs*. Englewood, CO: Libraries Unlimited.

Panzer, M., & Zeng, M. (2009). Modeling classification systems in SKOS: Some challenges and best-practice recommendations. Paper presented at the International Conference on Dublin Core and Metadata Applications (DC-2009), Seoul, South Korea. Available at: <http://www.nla.gov.au/lis/stdnrd/grps/acoc/documents/EPC132-38.1.doc>

People and Discoveries (1998). Available at: <http://www.pbs.org/wgbh/aso/databank/entries/dh23eu.html>

Ranganathan, S.R. (1967). *Prolegomena to library classification* (3rd ed.). Bombay, India: Asia Publishing House.

Rorty, R. (1982). *Consequences of pragmatism*. Minneapolis, MN: University of Minnesota Press.

Rorty, R. (1999). *Philosophy and social hope*. New York, NY: Penguin.

Tennis, J.T. (2002). Subject ontogeny: subject access through time and the dimensionality of classification. In *Challenges in knowledge representation and organization for the 21st century: Integration of knowledge across boundaries: Proceedings of the Seventh International ISKO Conference*. (Granada, Spain, July 10-13, 2002). *Advances in Knowledge Organization*, Vol. 8 (pp. 54-59). Würzburg: Ergon.

Tennis, J.T. (2007). Scheme versioning in the semantic web. *Cataloging and Classification Quarterly*, 43, 85-104.

Tennis, J.T., & Sutton, S.A. (2008). Extending the Simple Knowledge Organization System (SKOS) for concept management in vocabulary development applications. *Journal of the American Society for Information Science and Technology*, 59, 25-37.

Tennis, J.T., Thornton, K., & Filer, A. (2012). Some temporal aspects of indexing and classification: Toward a metrics for measuring scheme change. In *Proceedings of the 2012 iConference* (pp. 311-316). ACM Digital Library.

- Thornton, K., & Tennis, J.T. (2010, December). Eugenics and anatomy: Case study of change and stasis in classificatory structures. University of Washington, iSchool Research Fair.
- Tuominen, J., Laurene, N., & Hyvönen, E. (2011). Biological names and taxonomies on the semantic web—Managing the change in scientific conception. In G. Antoniou, M. Grobelnik, E. Paslaru Bontas Simperl, B. Parsia, D. Plexousakis, P. De Leenheer, & J.Z. Pan (Eds.), Proceedings of the Extended Semantic Web Conference (pp. 255–269). Lecture Notes in Computer Science, Vol. 6644. Berlin, Germany: Springer.
- Vickery, B. (1960). *Faceted classification: A guide to construction and use of special schemes*. London, UK: Aslib.
- Vickery, B. (1966). Faceted classification schemes. In S. Artandi (Series Ed.), *Rutgers Series on Systems for the Intellectual Organization of Information* (Vol. 5). New Brunswick, NJ: GLS Rutgers.