

Structural (In)visibility: Possible Effects of Constructing a Controlled Vocabulary in a Niche Domain

Audrey Lorberfeld

University of Washington iSchool
aulorbe@gmail.com

Elan May Rinck

University of Washington iSchool
evmay@uw.edu

ABSTRACT

The paper explores the challenges and implications that arose during the construction of a controlled vocabulary in an emerging, non-structured domain. It discusses the difficulty of defining a domain with evolving and intersectional boundaries, and how Wittgenstein's philosophy of language may aid designers in properly representing a domain's hierarchy when multiple language games are present. It also discusses adaptation of standard construction procedures to maintain domain representation. The analysis focuses on the potential effects that constructing such a controlled vocabulary could cause in the future evolution of that domain.

1.0 CONSTRUCTING A CONTROLLED VOCABULARY IN A NON-STRUCTURED DOMAIN

Controlled vocabularies seek to make natural language more artificial (Svenonius, 2003, 823). Further, through the standardization and disambiguation of terms used in everyday speech and published literature, controlled vocabularies aid in the general understanding of a specialized community and define semantic relationships between terms (Aitchison et al, 2000, 1). The particular type of controlled vocabulary discussed here is a thesaurus. In the larger context of controlled vocabularies, the purpose of thesauri is to increase the precision and recall of relevant information by providing a diverse set of words with particular focus on synonyms and broader terms (Aitchison et al, 2000, 2). This paper will explore the possible implications of constructing a thesaurus that aims to control the vocabulary of an emerging, non-structured domain. These possible implications stem from the act of creating a prototype thesaurus called the DIY Biohacking Thesaurus (DBT). The domain of the DBT is do-it-yourself (DIY) biohacking, a sub-community within

the larger world of biohacking. Biohacking is the activity of leveraging biological structures and technology outside of mainstream scientific norms to alter organisms and their functions. Similarly, DIY biohacking primarily focuses on applying these methods to the individual's body using widely available tools and materials.

The world of DIY biohacking resists classification and represents a non-structured domain. It pulls from multiple epistemologies concurrently, making the act of defining domain boundaries more difficult than usual. These epistemologies approximate a cross-section of biology, philosophy, and political science. DIY biohackers and their specialized knowledge operate outside the bounds of more socially acceptable, traditional domains. The domain lacks established ways of defining authoritative knowledge, such as peer review processes, which renders traditional domain analysis and term selection techniques potentially less effective. For this reason, we consider DIY biohacking an emerging, non-structured domain. These factors lead to potential ethical and operational implications in thesaurus construction for not only its creators and its users, but also for the domain itself.

Because of the complicated nature of classifying non-structured domains, it could be argued that attempting to classify them is a Sisyphean task. We believe, though, that while some of this criticism has merit, unstructured domains can benefit from classification, and the challenges they present are also reasons to develop controlled vocabularies. Additionally, in a domain without firm consensus of authoritative sources, precision in information classification and retrieval could be crucial to domain development. In the case of DIY biohacking, a controlled vocabulary can also enable domain members to quickly find accurate information about health concerns, safety best-practices, and appropriate materials for self-experiments. The benefits of such use cases would be measurable and important.

2.0 GENERAL ISSUES IN THESAURUS CONSTRUCTION

In analyzing the cost and benefit of thesaurus construction in emerging, non-structured domains, it is important to acknowledge that knowledge organization systems (KOSs) are widely accepted as biased, even in

traditionally structured domains. Hodge (2000) lists three critical characteristics of all KOSs:

1. The KOS imposes a particular view of the world on a collection and the items in it.
 2. The same entity can be characterized in different ways, depending on the KOS that is used.
 3. There must be sufficient commonality between the concept expressed in a KOS and the real-world object to which that concept refers that a knowledgeable person could apply the system with reasonable reliability. Likewise, a person seeking relevant material by using a KOS must be able to connect his or her concept with its representation in the system. (para. 8)
- KOSs, by definition, are both subjective and faithfully representative.

Thesauri fall prey to the same contradiction. The “aboutness,” or meaning, of an information object is inherently open to interpretation by both thesaurus constructor and user. Additionally, controlled vocabularies need to contend with the subjective nature of language. As Woolwine et al. (2011) note:

... indexing terms themselves do not have stable meanings. The meanings of indexing terms change as at least two contexts change: the context of the index itself (a shift in the meaning of one term can lead internally to shifts in meanings of other terms) and the context of the academic discourse to which indexing terms are meant to provide access. (p. 80)

Consequently, there arise two opportunities for inaccuracy to develop: semantic change in the organization structure itself and conceptual change within the domain.

Domains, by their very nature, are fluid; discourse continues and semantic meanings change. To a certain extent, this is a risk constructors of indexing languages acknowledge; the organizational payoff is worth minor semantic inaccuracies, if the language can serve its users effectively. Ambiguity in information organization could be acceptable, if the use of a controlled vocabulary truly meets users’ retrieval requirements (Svenonius, 2003, p. 837). When approaching potential pitfalls in the design of controlled vocabularies, user warrant, the methodology of culling terms from a domain population’s casual language, is of the utmost importance, and can easily outweigh concerns about domain representation.

Though developed to classify an emerging and non-structured domain, the DBT conforms to *ANSI/NISO Z39.15-2005 (R2010) Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies* (2005). We relied heavily on Aitchison,

Gilchrist, and Bawden’s *Thesaurus Construction and Use: a Practical Manual* (2000) and Soergel’s *Indexing Languages and Thesauri: Construction and Maintenance* (1974). As such, many of the design and construction decisions were informed by standards and guidelines intended to be applied to thesauri in more traditional, structured domains. In such cases, these guidelines represent best practices and the most effective way to create an accurate and useful controlled vocabulary. They are tried and tested in traditional domains and are agreed to be relatively effective (Batley, 2014; Clark & Aitchison, 2013). However, it is necessary to explore if these methods represent best practices when applied to emerging and non-structured domains.

3.0 REASONS TO DEVELOP A CONTROLLED VOCABULARY IN AN UNSTRUCTURED DOMAIN

We believe that the creation of controlled vocabularies in non-structured, emerging domains is worth pursuing because it allows exploration of the limits of classification. As Hjørland (2002) notes:

Most research on classification is about universal classification schemes and little has been done about special domains. The lack of interest in special domains is problematic for several reasons, one being that lack of knowledge on the classification of a special field also is reflected in the updating and revision of universal schemes. (p. 425)

In the process of classifying a non-structured domain, we can further explore the applicability of universal schemes and existing best practices. Limits of current classification can be investigated, and designers can better understand the flexibility of current best practice. However, because of the ethical concerns involved in classification, there must also be a practical reason for the development of a controlled vocabulary within a non-structured domain.

Development of controlled vocabularies is predicated on meeting the information retrieval needs of users. If successful information retrieval depends on it, a controlled vocabulary is likely necessary. However, Svenonius (2003) brings up additional considerations:

If a subject discipline is such that its writers tend to give their works noninformative titles, a CV is needed. A CV is needed if the vocabulary of the discipline exhibits very little representational predictability. On the other hand, if the discipline lacks a special terminology or if its vocabulary exhibits a great deal of linguistic indeterminateness, a CV may be less valuable to the extent that control is difficult to impose. (p. 837)

DIY biohacking both exhibits little representational predictability and a vocabulary with a high degree of linguistic indeterminateness. Thus, there is potential value in the construction of a controlled vocabulary for the domain, but there is also the potential that value could diminish due to the difficulty of developing control. So, how does a designer decide whether the development of a controlled vocabulary is justified?

One consideration is collocation. Even though the domain is informal, the scientific and research-oriented nature of DIY biohacking means that the collocation afforded by a controlled vocabulary could be a considerable boon to users. Increased precision in retrieval is also a convincing argument for the construction of a controlled vocabulary in this case. Because of the complex semantic landscape, information retrieval within this domain can benefit from the predictability of representation that a controlled vocabulary can provide. In short, the decision to develop a controlled vocabulary in a non-structured domain should be based on the value to the user in information retrieval, not the benefit of imposing structure itself.

4.0 UNDERSTANDING EPISTEMOLOGY AND SEMANTIC DYNAMICS OF UNSTRUCTURED DOMAINS

In the case of the DBT, once the user-benefit of controlled vocabulary construction was established, we had to address whether the dynamics within the domain allowed for successful classification. We found that the intersectionality of the domain required the development of a paradigm through which to parse the conceptual interactions and evolution of terms. Exploring the epistemological and semantic dynamics of the domain allowed us to determine how to best approach classification and what challenges could arise.

To best understand the epistemological issues within a non-structured domain, Wittgenstein's idea of language games and its surrounding scholarship is helpful. Wittgenstein's philosophy of language focuses on the disconnection of meaning from terminology and the emergence of meaning through term use. He notes that indeterminacy of language is not a result of irrational use, but of language being taken out of a normal context and applied elsewhere (Wittgenstein & Nyman, 1991). This is an important concept in the epistemological analysis of emerging and non-structured domains, where much of the lexicon is borrowed from other, bordering domains. This idea also presents a challenge to controlled vocabulary designers.

Wittgenstein's language games can further be used as a framework to explore a domain's semantic dynamics. Inherent in Wittgenstein's concept of linguistic indeterminacy is contextual semantic evolution (Wittgenstein & Nyman, 1991). As use changes, so do the semantic relationships between terms. This concept is

particularly relevant in a domain such as DIY biohacking, where terms from two distinct epistemologies can develop intimate relationships, such as "transhuman" and "RFID." Being aware of the increased presence of these phenomena in a domain such as DIY biohacking allowed us to approach domain analysis with an idea of what would need to be adapted.

5.0 DOMAIN DEFINITION AND ANALYSIS

A challenge present in constructing the DBT was appropriately defining an area of knowledge that may not have clear or stable boundaries while faithfully representing its language games. Domain fluidity can be a particular obstacle when thesaurus designers attempt to apply standard construction methods to non-structured domains, such as DIY biohacking. This problem stems from the intersectional nature of such a domain and the accompanying lack of agreement within its discourse community about what constitutes an authoritative source. In fact, there is no clear consensus as to whether there should be an authority-hierarchy at all. This necessitates serious consideration of the benefit and cost of the semantic organization of these information bases. In such cases, successful and thorough domain analysis is even more crucial.

Tennis's (2003) work "Two Axes of Domains for Domain Analysis" and Hjørland's (2002) work "Domain Analysis in Information Science: Eleven Approaches--Traditional as well as Innovative" provide cogent and adaptable paradigms through which to examine non-standard domains like biohacking. As a transmutation of biology, philosophy, and political science, biohacking is similar to Biomedical Ethics, which Tennis describes as pulling membership from many freestanding domains (2003, 194). In such domains, the implementation of Tennis's proposed areas of modulation is more difficult, because the body of knowledge incorporates multiple knowledge bases. This does not mean domain analysis is impossible, but it does mean that Tennis's second axis, degrees of specialization, becomes even more crucial. This axis allows designers to effectively define a subdomain, even across areas of modulation.

Analyzing degrees of specialization allows for a clear definition of scope by accounting for focus and workable domain boundaries that begin to emerge at varying levels of specificity, such as commercial biohacking, applied transhumanism, anarchist biohacking (or grinding), and DIY biohacking. As designers, identifying a degree of specialization that provided clear domain boundaries, but that was not so granular that it would fail to provide significant value for information discovery, was paramount. The domain of DIY biohacking possesses enough intersectionality that a controlled vocabulary could provide value to a significant portion of the domain population, but is also granular enough to avoid an unwieldy domain scope.

In order to parse the various epistemological and terminological threads present within DIY biohacking, we turned to Hjørland's eleven approaches to domain analysis (2002). Hjørland's work allowed us to establish an appropriate "domain analysis paradigm" with the acknowledged caveat that domains that are comprised of mostly unstructured sources and a loosely organized discourse community do not lend themselves to many of his analysis approaches. Namely, traditional literature guides, empirical user studies, bibliometric studies, and studies of institutional structure is difficult to do in a traditional fashion to such domains. Instead, our domain analysis first focused on limited terminological studies, then on epistemological and historical studies in order to identify the domain's paradigms. With this background, we were then able to effectively implement and adapt Hjørland's approaches of literature guides and document study.

Though not implicit in Hjørland's work, we believe the order of implementation in domain analysis is especially important in non-structured domains, because purposeful order allows the designer to have an appropriate epistemological paradigm from which to analyze more granular features of the domain in addition to allowing the designer to account for unintentional bias. In a non-structured domain, establishing an understanding of domain-specific epistemology is particularly relevant. There is a risk of incorrectly designating terminology scope, if the designer does not have a full epistemological understanding.

Attempting thesaurus construction without a sound epistemological understanding of the domain at hand not only has the potential to decrease the accuracy of information retrieval, but could also have the unintended effect of shaping an evolving domain and imposing non-native structures. This raises questions about the warrant of domain definition decisions and whether the controlled vocabulary designers have a responsibility to faithfully represent realistic domain boundaries. One might argue such a decision is acceptable, if necessitated by improved information retrieval, but we do not believe it is so simple. While establishing a degree of specialization is intended to overcome ambiguity in the domain, it also creates boundaries that are structural, rather than organic. These boundaries are a product of the needs of classification and information retrieval, not necessarily the needs of the domain as a whole. It is possible that construction decisions could have lasting impacts on the social, intellectual, and epistemological evolution of the domain, if such a thesaurus was used by domain members.

6.0 CONSTRUCTION AND SEMANTIC INDETERMINACY

Approaching DIY biohacking through the critical scholarship of Tennis and Hjørland resulted in a decision

to manually select terms and, due to the emerging nature of the domain, prioritize user warrant for preferred term selection. Much like domain analysis, however, the accepted methods for term selection can be difficult in an emerging, non-structured domain. NISO (2005) recommends three initial approaches: an expert committee approach, an empirical approach, or a combined approach (Z.39-19-2005 11.1.3). While the committee approach is an excellent way to construct a vocabulary in structured domains, it can be problematic in non-structured discourse communities. In a domain like DIY biohacking, identifying experts within the community is difficult, even within the limited scope of the a combined approach. NISO's (2005) recommendations for empirical analysis are more helpful for a non-structured domain, but still require curated adaptation (Z.39-19-2005 11.1.3.2). The standards offer two options for empirical analysis: inductive and deductive. Deductive analysis allows the designers to begin with the domain paradigm discovered through earlier research, thereby honoring a domains' natural dynamics and structure. If employing inductive analysis, on the other hand, the designers create and impose a hierarchical structure during domain analysis and term selection. Essentially, the designer is performing a structural analysis of the domain concurrently with a literature and terminological study. We believe this could present practical issues in a non-structured domain because of the increased rate of polysemy.

Polysemy is not an issue specific to emergent or non-structured domains, but a variety of equal and competing language games may be. In terms of divergent domain language games, our earlier application of Wittgenstein's theories raises several more potential challenges in this context. Increase polysemy was the first difficulty presented by the domain's high level of language indeterminacy. From a pragmatic standpoint, this should not dissuade designers from attempting to classify non-structured domains, but does necessitate careful observance of the semantic dynamics within them.

As explicated in *ANSI/NISO Z39.19-2005* (2005), "the need for vocabulary control arises from two basic features of natural language . . . two or more words or terms . . . used to represent a single concept" and "two or more words that have the same spelling [that] represent different concepts." The issue of polysemy and homonymy are universal challenges in the construction of controlled vocabularies. Standard practice designates terminology study and selection as somewhat outside of epistemological studies, but Wittgenstein's philosophy may indicate this is not the best way to approach domains with more lexical indeterminacy (Wittgenstein & Nyman, 1991).

As Pilerot (2012) notes, "A more thorough understanding of a concept can accordingly be reached by taking into

consideration the concept's systematic connections to the context in which it is aimed to function" (561). This underscores the importance of looking at a term's epistemological origin when analyzing domains with high levels of discourse-community cross-over. We found that a high prevalence of language indeterminacy can lead to a magnification of existing issues in thesaurus construction.

A secondary issue that arose from our analysis through the lens of Wittgenstein was the prevalence of cross-hierarchical relationships between terms due to the complex semantic dynamics of the domain. This issue is not one that should dissuade designers from constructing a controlled vocabulary, but it does require careful analysis and decision-making. The very nature of an indexing language allows the constructor to confront this issue from a pragmatic perspective through the use of equivalence relationship designators. But how do we best overcome this issue from an epistemological perspective? Special attention to the language games of the domain can help.

As Andersen and Christensen (1999) state, "the meaning and purpose of a document is not a property inherent to it. Rather its linguistic and conceptual meaning is determined by external factors, within the framework of the language game A document cannot define itself" (p. 16). In any domain, it is crucial for the designer to analyze the accuracy of the indexing language as a representation of the semantic and terminological features of the discourse community in question. By considering, investigating, and appropriately applying a domain's language games during construction, designers can potentially address a domain's less-standard features, such as frequent polysemy caused by language indeterminacy and complex semantic relationships between terms.

Structured domains, such as classical philosophy, experience similar issues. Such domains' conceptual and semantic maps are largely based on ordinary language, in which terms' meanings frequently evolve while also retaining their former definitions. Thus, a term has a variety of meanings that are largely contextual, but all crucial in the semantic representation of the domain. Such a phenomenon supports Mazzochi and Tiberi's (2009) assertion that:

. . . the cognitive authority that stipulates the basic rules for the use of any term in a given knowledge field, and thus its meaning, does not reside in the documents as such, but rather in the accepted paradigms (here intended in the broadest sense) of the field, on the basis of which the documents themselves are produced. (p. 108)

DIY biohacking shares many features with the philosophy domain explored by Mazzochi and Tiberi. Its root is philosophical and many terms within the domain are dependent on their connection to varying theoretical origins. Moreover, there is a high rate of semantic transformation and an associated prevalence of polysemy. In order to faithfully represent the domain, a thesaurus of DIY biohacking would "be required to cope with the complexity and the level of complexity and level of abstraction of the philosophical conceptual structures" (Mazzochi & Tiberi, 2009, p. 104). This is a task that can be accomplished once realized, but a designer must examine if it results in a true reflection of the domain's structure.

What DIY biohacking does not share with philosophy is a singular emphasis on semantic abstraction. DIY biohacking is a multivariate domain with roots in philosophy, biology, and political science. This epistemological amalgamation creates atypical relationships, with some terms building meaning with time, others having their meaning replaced entirely as new discoveries are made, and still more falling prey to semantic obsolescence. In essence, DIY biohacking is subject to a few separate language games: those which drive the semantic representation of philosophical ideas, those which drive the semantic representation of scientific ideas, and those which govern the interaction between the two. It is necessary to explore how to best address and effectively honor multiple language games in the structure of a controlled vocabulary.

7.0 IMPLICATIONS OF CONTROLLED VOCABULARY CONSTRUCTION

On the one hand, such complex domains can provide thesaurus designers a veritable playground for classification thought-experiments and opportunities for critical inquiry into traditional construction methodologies. On the other hand, non-traditional domains raise ethical concerns that constructors cannot ignore. In addition to questions about the practical possibilities of classification, designers must also address concerns related to cultural intrusion, commodification, and colonization.

The constructor wields considerable power over a domain through classification. In his exploration of the mind's role in information, Blair (2006) states, "the selection of information on the system embodies... the discrimination of the individuals who designed the system" (p. 284). Indeed, any system which filters information reflects the designer's motives, rather than reality. This is a concept that has been explored extensively in traditional controlled vocabulary construction (Marshall, 1977; Drabinski, 2013; Fleischmann et al., 2009; Olson & Schlegl, 2001).

This inherent bias may not always be obvious to the system's user, especially in domains that do not have a historical use of controlled vocabularies, like DIY biohacking. When a hierarchical structure is gleaned from a domain that does not present one naturally, users may not be aware that the structure of the controlled vocabulary is dependent on decisions that were made by designers who are not domain experts. This structural invisibility may be a real concern in non-structured, emerging domains that have native structures that are still evolving.

Mai (2008) writes that "the decisions a designer of controlled vocabularies needs to make . . . are informed by the designer's epistemological stance" (p. 17). As designers without expert domain knowledge, the facets within the DBT are primarily informed by the Classification Research Group's Universal Facets: Ideas, Entities & Objects, and Actions (Aitchison, Gilchrist, & Bawden, 2000). This hierarchical structure, while effective, reflects an information science epistemology, rather than one native to the domain. Implementing an artificial structure on a domain to control term relationships blurs the epistemological nuances of the chimerical relationships that might naturally occur. For instance, some within the domain might think that relating "biohacker" to "implant," rather than to "maker," a near synonym, would be most appropriate in order to reflect the intimate relationship between person and technology.

Structurally, this is a relatively easy challenge to overcome. It is important to maintain awareness of the increased semantic transformation that occurs in emerging, non-structured domains. Designers can use contemporary construction techniques as guides when approaching this challenge. Olson (2007) writes that

Implicit in the RTs [related terms] is a recognition that hierarchy alone is insufficient . . . this relationship covers associations between terms that are neither equivalent nor hierarchical, yet the terms are semantically or conceptually associated to such an extent that the link between them should be made explicit in the controlled vocabulary, on the grounds that it may suggest additional terms for us in indexing and retrieval . . . in the environment of information retrieval and its potential for sophisticated change, the associative relationship remains a catchall. (p. 516)

The related term (RT) designator allows the flexibility to represent cross-epistemological relationships that may not be well-suited for strict hierarchical organization. While the structure of the DBT reflects the (in)visibility of its constructors, the designated inter-facet relationships allow a relatively faithful representation of the epistemologies

and vocabulary of the domain. The use of extensive inter-facet relationship designators attempts to curb the potential epistemological rearrangement a facet structure might impose. However, the threat of colonizing another population's information world, and, consequently, limiting the growth of an emerging domain, is still present.

7. COLONIZATION OF INFORMATION IN CONTROLLED VOCABULARIES

Tackling the pragmatic issues of structural representation does not alleviate the designer's responsibility to domain members. Martinez and Guimaraes (2013) write that a sure way to eradicate a certain group or people from history is not to name it at all (p. 22). The role of the controlled vocabulary designer, then, is increasingly complicated. Not only are designers supposed to acknowledge (maybe even embrace) epistemological bias and strive, as much as possible, for objective reliability, they are also tasked with validating and preserving the existence of ideas, people, organizations, and objects through language.

It is clear that the user's domain interpretation is deeply affected by both the structure and term selection of the controlled vocabulary used to access that domain. Unruh (1980) writes, "social worlds originate in the lexicon of everyday life" (p. 275). Controlled vocabularies can shift that lexicon by prioritizing terms or concepts over retrieval precision. This serves an intentional purpose, but it can also affect the domain's zeitgeist by encouraging prescribed language use and erasing or maintaining ideologies. As Wittgenstein writes, "imagine that the language . . . was the whole language of a tribe. The children are brought up to perform these actions . . . to use these words as they do, and to react in this way to the words of others" (quoted by Blair, 2006, 77). If designers accept Wittgenstein's theory of language games, it becomes clear that prescribing language use deeply affects the ways that individuals and cultures approach the knowledge associated with that language. Indeed, prescribed language can shift the actions and growth of a domain.

Thus, the importance of reflecting a domain's epistemologies becomes even more apparent when applying a controlled vocabulary to a domain that is still emerging and developing. Hartel (2003) writes that "information plays a critical role in hobby social worlds. The lack of any centralized bureaucracy causes a dependence on mediated communication, namely: books, magazines, chat rooms, newspapers, and various other information forms" (p. 231). The increasing omniscience of the computer, specifically the ease of access to information that the Internet provides, increases the possible effects of a controlled vocabulary in an emerging domain. The creation of a thesaurus to aid in the serendipitous information exploration available to anyone

with a computer and an Internet connection is an incredible tool, but its wide accessibility can also exacerbate the aforementioned issues surrounding prescribed language usage.

One of the greatest tools to avoid colonization of domain information is community input. When asked if they thought having a thesaurus would benefit people trying to get more information about the world of DIY biohacking, Lydia Fazio and Richard Ryan, prominent members of the New York biohacking community, identified the value, but raised concerns. Fazio stated, "a unified language is necessary for the growth of the movement, but the language may also strangle the movement" (personal communication, September 26, 2015). By providing a controlled vocabulary, access is facilitated, but so is semantic structure. In a domain where pushing the bounds of current thought is encouraged, access to information and control must be balanced.

Fazio and Ryan also recognized the basic challenges of the domain being non-structured: there are no vetted guidelines, there is no mission statement, no centralized wiki, no one-stop-shop for people outside the domain looking to gain information. They felt that a controlled vocabulary, if implemented, would be useful for information retrieval for a wide base of users. Interestingly, they also saw value in the hierarchical structure itself, noting that it could provide a shared point of reference and description among different domain factions and may increase unity of thought within the domain by explicitly pointing to the domain's philosophical roots (Fazio and Ryan, personal communication, September 26, 2015). This represents a level of impact that controlled vocabularies may not have on structured and established domains. The implementation of a controlled vocabulary on a domain such as DIY biohacking enables the possibility that its creation may enforce a structure that, while epistemologically sound, was not previously recognized in the discourse community.

Ryan saw this potential impact as both positive and negative. He noted that, while the creation of such a standardized, controlled vocabulary might help biohackers represent the field more accurately, standardizing the language within the domain could also allow outside influences, such as corporate interests and the government, to easily exploit the community and draft potentially hazardous, regulatory legislation (Ryan, personal communication, September 26, 2015). Those possibilities present an ethical dilemma for designers and pushed us to consider our own intentions and epistemological biases in the construction of the DBT.

8. FURTHER RESEARCH

While the construction of the DBT has allowed the exploration of a number of issues in classification, it has

also highlighted phenomena that merit further investigation. Technology provides some potential solutions to issues discussed in this paper. In particular, the integration of folksonomies and a standard controlled vocabulary, as explored by Woolwine et al. (2011), may provide a partial solution to some concerns related to the epistemological co-option of a domain by designers. User-based term-validation or tagging items with controlled terms, such as the ARTigo project (2015), should also be explored further. If properly executed, this could allow domain members to have more immediate and ongoing input in defining and refining term relationships, leading to a more accurate epistemological reflection of a contemporary domain.

The construction of the DBT also raises questions regarding the progression and growth of a domain. At what point is a domain's structure mostly fixed? How fixed must a domain's structure be to support successful and faithfully representative classification? The investigation of these issues may provide further guidance in adequately classifying the knowledge in emerging, non-structured domains.

Finally, our work on the DBT has underscored the importance of performing classification work in emerging, non-structured domains, as it allowed us to parse the limits of universal classification. By calling for continued research in this area, we hope to contribute to the refining of classification theory in order to not only improve standard classification practices, but also to normalize and facilitate the classification of non-structured domains.

REFERENCES

- Aitchison, J., Gilchrist, A., & Bawden, D.. (2000). *Thesaurus construction and use: A practical manual* (4th ed.). London, England: Aslib IMI.
- Anderson, J., & Christensen, F.S. (1999). *Wittgenstein and Indexing Theory*. Paper presented at 10th ASIG SIG/CR Classification Research Workshop. Washington, DC: ASIG.
- ARTigo. (2015). About. Retrieved October 31, 2015, from <http://www.artigo.org/about.html>
- Batley, S. (2014). *Classification in theory and practice* (2nd ed.). Oxford, England: Chandos Publishing.
- Blair, D. (2006). "Wittgenstein's philosophy of language and mind." *Wittgenstein, language and information* (pp. 27–267). Dordrecht, Netherlands, Springer.
- Clarke, S. D., & Aitchison, J. (2013). The thesaurus: A historical viewpoint, with a look to the future. *The thesaurus: Review, renaissance, and revision*, 37(3-4), 5-22.

- Drabinski, E. (2013). Queering the catalog: Queer theory and the politics of correction. *The Library Quarterly*, 83(2), 94–111.
- Eggleson, K. (2014). Transatlantic divergences in citizen science ethics: Comparative analysis of the DIYbio code of ethics drafts of 2011. *NanoEthics*, 8(2), 187–192.
- Fleischmann, K. R., Oard, D. W., Cheng, A., Wang, P., & Ishita, E. (2009). Automatic classification of human values: Applying computational thinking to information ethics. *Proceedings of the American Society for Information Science and Technology*, 46(1), 1–4.
- Hartel, J. (2003). The serious leisure frontier in library and information science: Hobby domains. *Knowledge Organization*, 30(3/4), 228–238.
- Hjørland, B. (2002). Domain analysis in information science; Eleven approaches--Traditional as well as innovative. *Journal of Documentation*, 58(4), 422–462.
- Hjørland, B. (2004). "Domain Analysis in Information Science." In *Encyclopedia of Library and Information Science*. New York, NY: Marcel Dekker.
- Hodge, G. (2000). Systems of knowledge organization for digital libraries: Beyond traditional authority files. Retrieved from <http://www.clir.org/pubs/reports/pub91/contents.html>.
- Hudon, M. (1999). An assessment of the usefulness of standardized definitions in a thesaurus through interindexer terminological consistency measurements. *Dissertation Abstracts International, Section A: The Humanities and Social Sciences*, 60(1), 10.
- Love, K., & Pole, C. (2012). *Ethics in social research*. Bradford: Emerald Group Publishing Limited.
- Marshall, J. (1977). *On equal terms: A thesaurus for nonsexist indexing and cataloging*. New York, NY: Neal-Schuman.
- Mai, J. (2008). Actors, domains, and constraints in the design and construction of controlled vocabularies. *Knowledge Organization*, 35(1), 16–29.
- Mai, J. (2013). Ethics, values and morality in contemporary library classifications. *Knowledge Organization*, 40(4), 242–253.
- Marchese, C., & Smiraglia, R. (2013). Boundary objects: CWA, an HR firm, and emergent vocabulary. *Knowledge Organization*, 40(4), 254–259.
- Martinez Ávila, D., & Guimarães, J. (2013). Library classifications criticisms: Universality, poststructuralism and ethics. *Scire*, 9(2), 21–26.
- Maso-Maresma, G., & Sebastia-Salat, M. (2013). The integration of folksonomies within a thesaurus in a social science web portal: SIDBRINT. *Information Research*, 18(3).
- Mazzocchi, F., & Tiberi, M. (2009). Knowledge organization in the philosophical domain: Dealing with polysemy in thesaurus building. *Knowledge Organization*, 36(2-3), 103–112.
- Meyer, M. (2013). Tinkering initiative biology: Policies and issues of do-it-yourself biology. *Medicine Sciences: M/S*, 29(2), 36–37.
- NISO. (2005). *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. ANSI/NISO Z39.19-2005.
- Olson, H. (2008). How we construct subjects: A feminist analysis. *Library Trends*, 56(2), 509–541.
- Olson, H., & Schlegl, R. (2001). Standardization, objectivity, and user focus: A meta-analysis of subject access critiques. *Cataloging & Classification Quarterly*, 32(2), 61–80.
- Pilerot, O. (2012). LIS research on information sharing activities--People, places, or information. *Journal of Documentation*, 68(4), 559–581.
- Smiraglia, R. (2014). *Elements of knowledge organization*. Cham: Springer International Publishing.
- Soergel, Dagobert. (1974). *Indexing Languages and Thesauri: Construction and Maintenance*. Los Angeles: Melville.
- Svenonius, E. (2003). "Design of controlled vocabularies". In A. Kent & H. Lancour, *Encyclopedia of library and information science* (pp.822-838). New York, NY: Marcel Dekker.
- Tennis, J. (2003). Two axes of domains for domain analysis. *Knowledge Organization*, 30(3/4), 191–195.
- Vayena, E., & Tasioulas, J. (2015). "We the scientists": a human right to citizen science. *Philosophy & Technology*, 28(3), 479–485.
- Wilson, P. (1968). *Two kinds of power: an essay on bibliographical control*. Berkeley: University of California Press.
- Wittgenstein, L., & Nyman, H. (1991). *Philosophy. Synthese*, 87(1), 3–22.
- Woolwine, D., Ferguson, M., Joly, E., Pickup, D., & Udma, C. (2011). Folksonomies, social tagging and scholarly articles. *Canadian Journal of Information and Library Science*, 35(1), 77–92.