Sources of Change for Modern Knowledge Organization Systems

Michael Lauruhn and Paul Groth

Elsevier Labs, 1600 John F. Kennedy Boulevard, Suite 1800, Philadelphia, PA,
<{m.lauruhn, p.groth}@elsevier.com>

Mike Lauruhn is a librarian, working as Research Technology Director at Elsevier Labs. His current research areas include linked data, taxonomies and ontologies, mark-up and annotation, research data lifecycles, and other issues affecting research communications. He is currently a member of the Dublin Core Metadata Initiative's Governing Board. Before joining Labs in 2010, he held consulting and technical positions helping large companies and organizations define and implement taxonomies and metadata schemas. Mike's earlier work experience includes cataloging for the California Newspaper Project at the Center for Bibliographic Studies and Research. Mike received his MLS from UCLA.

Paul Groth is the Disruptive Technology Director at Elsevier Labs. He holds a PhD in computer science from the University of Southampton (2007) and has done research at the University of Southern California and the Vrije Universiteit Amsterdam. His research focuses on dealing with large amounts of diverse contextualized knowledge with a particular focus on the web and science applications. This includes research in data provenance, data science, data integration, and knowledge sharing. He is co-author of two books as well as numerous academic articles. He blogs at http://thinklinks.wordpress.com. You can find him on twitter: @pgroth.

Lauruhn, Michael and Paul Groth. 2016. "Sources of Change for Modern Knowledge Organization Systems." *Knowledge Organization* 43(8): 622-629. 27 references.

Abstract: Knowledge organization systems (KOSs, e.g., taxonomies and ontologies) continue to contribute benefits in the design of information systems by providing a shared conceptual underpinning for developers, users, and automated systems. However, the standard mechanisms for the management of KOS changes are inadequate for systems built on top of thousands of data sources or with the involvement of hundreds of individuals. In this work, we review standard sources of change for KOSs such as institutional shifts, standards cycles, cultural and political changes, distribution, etc., and then proceed to catalog new sources of change for KOSs ranging from massively cooperative development to always-on automated extraction systems. Finally, we reflect on what this means for the design and management of KOSs.

Received: 31 July 2016; Revised 16 October 2016; Accepted 27 October 2016

Keywords: change, KOSs, MeSH, processes, stakeholders, crowdsourcing

1.0 Introduction

Knowledge organization systems (KOSs) such as ontologies, terminologies, data dictionaries, and classification schemes provide the foundation for a variety of applications. These applications range from classification of objects, indexing processes, and traditional information retrieval (IR) systems to semantic web applications, question answering, and rule-based systems. While the core goal of many KOSs is to resolve entities and concepts for the applications they serve, newer functionality includes reasoning and discovery. Traditionally, KOSs have depended on manual processes that were largely akin to an editorial process based entirely on human supervision, recommendation, and decisions. These were typically done by, or with input and influence from, subject matter experts or collection experts. As KOS creation and maintenance processes evolve, one significant change is the level of human input that can effect change.

We note how sources of change differ as KOSs evolve across various models over time. We begin by looking at issues related to evaluating and correcting bias in early KOSs (*Library of Congress Subject Headings* and *Dewey Decimal Classification*) that were designed for library cataloging. These models had a wide breadth in scope and regularly scheduled distribution processes, initially exclusively in print. These earlier KOSs rely heavily on full-time editors and dedicated subject matters, and the development and dissemination process can be equated to an editorial system that produces editions on a regular schedule. Over time, KOSs with scopes that are more specific became more prevalent. In this paper, we will examine a pair that are created for indexing biomedical literature and databases. The editorial processes are still based on human decision making, but begin to come from a more distributed group of experts who are both users of and contributors to the model. Likewise, the sources of change also begin to become more application focused. As the contributor model becomes more distributed, the release of updates and change evolves into an ongoing process with more frequent versions that start to resemble a software release cycle.

In this work, our aim is to provide a catalog of sources of change of which designers and users of KOSs need to be aware. Figure 1 depicts the nine sources of change we call out in this paper and how they map into the major constitutes that impact the maintenance and creation of KOSs.

To substantiate this list, we begin by a review of the existing literature. We then proceed to catalog new sources of change with appropriate exemplars. This is followed by a discussion about the implications all these sources of change have on KOS design and management.

2.0 Sources of KOS change: the existing view

2.1 Change as an editorial process

One of the major drivers behind change in KOSs, particularly those that are largely-distributed across disciplines, is a desire to remain up to date in terms of cultural sensitivities and perceived bias. In his 1971 book, *Prejudices and Antipathies: A Tract on the LC Subject Heads Concerning People*, Sanford Berman (1971) outlined many of the subject areas in which the *Library of Congress Subject Headings* (*LCSH*) were insensitive and out of touch with the society that libraries operated in. These included areas such as race and ethnic groups, gender roles in society, politics, sexuality, and others. In the 1971 introduction, Berman cited previous literature, which identified earlier justifications for the bias that was prevalent in the *LCSH*. The author then encouraged others to join him in seeking

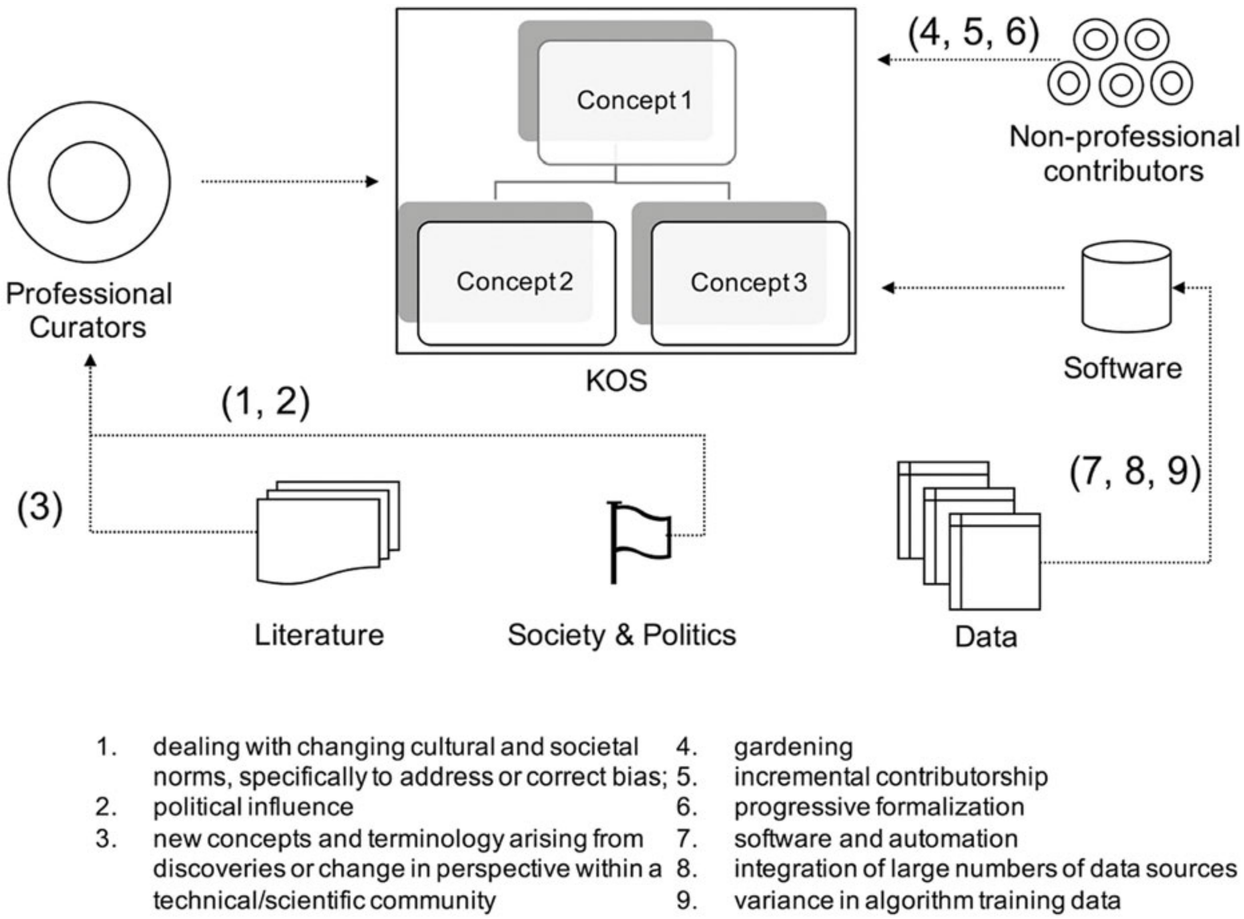

Figure 1. Sources of changed mapped to the major actors and entities within KOS creation and maintenance processes.

to "remedy long-standing mistakes and to gain for the profession a genuine, earned respect among people who read and think" (17).

In his subsequent Preface to the 1993 reprinting of *Prejudices and Antipathies*, Berman reflected on some of the changes that *LCSH* had undergone since the book was originally published (5): "But these changes [that had occurred], however welcome, are no cause for gloating. It took 13 years for LC to scrap JEWISH QUESTION and 18 to eliminate YELLOW PERIL, hardly examples of swift response and profound sensitivity."

Eleven years later, Knowlton (2004) published a study that compiled Berman's suggestions and tracked the changes that had occurred. It found that 39% of the subjects were changed in a manner that nearly matched Berman's suggestions, while 36% were unchanged, and 24% changed in a manner that partially captured the suggestions. Knowlton acknowledged that compilers of *LCSH* had addressed bias in a serious manner in the years following Berman's original work. He notes that many of the suggestions that went unchanged (128) "simply reflect a difference of opinion on the literary merit of subject headings changes" while elsewhere large swaths may reflect continuing bias.

To outside observers, it might seem that making changes in order to remain up to date on cultural and societal norms may make common sense and "not" making changes borders on egregious and insensitive behavior. However, even if an organization wants to make changes, often times those changes may be difficult to execute due to processes, rationale, and decisionmaking required to reach consensus and redefine structures in the model. These delays can be further extended when the KOS is traditionally released on a regular basis and there is a reluctance to make changes that may have to be changed again or changed back in the future. As an example, Green (2015) highlights many of the issues dealing with how indigenous peoples in the United States are represented in the *Dewey Decimal Classification* (*DDC*). Complexities include a many-to-many relationship that exists between ethnic groups and federally recognized tribes in the United States. In addition, while there may be a concept recognizing an ethnic group, that group is not represented as a sovereign nation. In addition, there may be some confusion as some nations are represented as geographic concepts. In acknowledging some of the claims of bias against the *DDC*, Green notes (220) that "some claims are perhaps based on misunderstanding, but some point to areas where the *DDC* can be improved." Green also illustrates how the decision-making process that leads to eventual change will include a vetting with the indigenous communities in question. This is an effort so that *DDC*'s classification principles and practices are upheld while being "true to the voice and perspective of the peoples being represented" (221).

Another source for change and, in some cases, influences "against" making desired changes to a KOS come from stakeholders not directly affiliated with the day-to-day maintenance of a KOS and their constituencies. In the winter of 2014, a student group from Dartmouth College began a movement to remove "illegal aliens" as a term in the *LCSH* (Qin 2016). The grassroots efforts of the Coalition for Immigration Reform, Equality and DREAMers (CoFIRED) resulted in a petition being presented to the Library of Congress in the summer of 2014. In 2016, the Library of Congress made clear its intentions to remove the term "illegal alien" from *LCSH* and replace it with "noncitizen" and "unauthorized immigration" (Aguilera 2016). This change was met with objections from members of the United States Congress, with members going so far as to add a clause to an Appropriations Bill (https://www.congress.gov/congressional-report/114th-congress/house-report/594/1) that instructs "the Library [of Congress] to maintain certain subject headings that reflect terminology used in title 8, United States Code," and subsequently leaving the terminology intact.[1]

As a means to create a larger community of contributors, the Library of Congress has formal channels to allow for a larger group of organizations and institutions to make proposals for changes in *LCSH*. The Subject Authority Cooperative (SACO) program uses the metaphor of the funnel to illustrate ideas and topics moving through a system of deliberation and professional judgement. Funnels are groups of libraries or catalogers that work in subject areas or specific regions and have joined together to contribute subject authority records for inclusion in the *LCSH*.

The examples above show widely-distributed, cross-disciplinary vocabularies that have been traditionally distributed in regular releases of volumes and editions, where it is not desirable to roll-back changes, thus significant changes are likely to be deliberated over time. It is worth noting that each of the previous examples deals with KOSs that include people, whether as ethnic groups, races, populations, or various other groups as part of their subject matter. The changes illustrated to the KOSs are based upon desires to correct previous biases that exist in the models.

2.2 Application-specific sources of change

Other KOSs have been more specific in their scope and faced different sources of change. Because of their specialized mandate, audience, and governance, they have different change processes. In many cases, the most common change for these KOSs is the addition of new concepts that are discovered within their domain. For the most part, these types of additions are not controversial and the processes depart from large interdisciplinary editorial-style deliberation to one where decision-making is left largely to dedicated subject matter experts.

Medical Subject Headings (MeSH) is the United States National Library of Medicine's (NLM) thesaurus. MeSH has section staff members who are responsible for ongoing revisions to the MeSH vocabulary. The MeSH website identifies three sources for changes: Subject specialists make changes in the areas of their expertise, indexers and others may make suggestions to the subject specialists, and the staff collect new concepts and terminology from scientific literature and emerging research (U.S. National Library of Medicine 2015a).

In describing their approach to decisions about changes to MeSH, the NLM (2015b) describes an approach that focuses on the primary use cases for the model, indexing and cataloging scientific literature. There seems to be a stronger emphasis on usefulness toward those tasks than there is for completeness from a strict ontological viewpoint:

> There are many factors that must be considered in deciding whether to add a MeSH descriptor. An interest in one species of a given genus, may lead to interest in some other species or even all of that genus. Yet, if there is little published about the other species, there is little purpose or advantage in creating a myriad of new descriptors in a vocabulary designed to describe the subject content of published literature. Before new descriptors are introduced, there is careful consideration of how the concept is currently indexed or cataloged. If the existing descriptors and qualifiers (subheadings) precisely characterize or identify the literature on the subject, there may not be a need for a new descriptor. Both too much change or too little change are to be avoided as MeSH is kept current with changes in biomedical knowledge.

In 2014, the NLM formed the Linked Data Infrastructure Working Group to investigate and determine best practices for publishing linked data (Bushman Anderson and Fu 2015). As part of the initiative, publishing MeSH in RDF was selected as a linked data pilot. The first beta version of MeSH RDF was based on the 2014 version of the vocabulary. When the next beta release of MeSH in RDF was released in June 2015, it was based on the 2015 version of MeSH. However, this marked a significant change, as moving forward NLM was able to make daily updates. MeSH in RDF has since moved out of beta and the daily update process is able to capture "emergency updates to MeSH." As an example, in Spring of 2016, descriptors for Zika Virus and Zika Virus Infections were added as Zika Virus epidemic spread in South America. Those changes will be incorporated in an annual version in the fall when the static graph for the 2016 version of MeSH is generated.

2.3 Distributed stakeholders

EMTREE is a proprietary thesaurus created by Elsevier in order to support indexing of EMBASE, the company's database of literature in the biomedical domain. At the beginning of 2016, the thesaurus contained more than 73,000 preferred terms and more than 310,000 synonyms (Elsevier 2016a). In "Change Management for Distributed Ontologies," Klein (2004) studied change processes for EMTREE and documented another largely human-directed process. Similar to MeSH, it was noted that most changes were to account for new terms and concepts that the EMTREE users suggested. Users indexing articles suggest additions throughout the year and accept that additions first go to a development version and later (between 3 and 15 months) appear in a production version. Also similar to MeSH, decisions to add concepts are based largely upon the frequency with which they appear in literature and that the two most common reasons for not including new concepts is that the information is either incorrect or does not occur often enough. EMTREE also contains mapping to MeSH, so another form a change that must be accounted in EMTREE is changes introduced in MeSH.

Another example from the sciences shows a more specific topic and a more distributed source of inputs from people. The Gene Ontology (GO) project is a bioinformatics project that builds and maintains ontologies for more than forty thousand biological concepts. The primary use of the ontology is to represent concepts used to annotate experiments that feature gene functions as reported in scientific articles and papers. The ontology is in a constant state of change to capture new discoveries (http://geneontology.org/page/about). Klein describes a process where a small number of full time curators work on the vocabulary and its relations but relies on GO users to make suggestions for new terms or edits. A change request system allows those users to track their submissions online and allows other users to see changes that are going through the submission process. As with MeSH and EMTREE, the primary source of changes is to add new terms, however many changes are also related to creating or updating relations within GO.

To summarize, we see the following sources of change documented in the literature:

1. Dealing with changing cultural and societal norms, specifically to address or correct bias;

2. Political influence; and,
3. New concepts and terminology arising from discoveries or change in perspective within a technical/scientific community.

3.0 New sources of change

The existing literature has focused primarily on sources of change stemming from practices associated with professional curators of KOSs. These include subject specialists, curators, or knowledge engineers. Two new actors, non-professional contributors (i.e., the crowd) and software, are now critical to the development of KOSs; whereas, in prior generations they were somewhat ancillary. The involvement of these actors fundamentally changes how KOSs are created and maintained.

3.1 Crowdsourcing

The emergence of Wikipedia and other crowdsource-based information systems has clearly impacted thinking behind the construction of KOSs. Vos (2006) described the differences between Wikipedia and Delicious categorizing systems and those of MeSH and the *DDC*. The bottom up style of the crowd sourced system is clearly evident and the overall network structure follows a stronger power-law like distribution.

Suchecki et al. (2012) studied the evolution of Wikipedia's category structure from its inception in 2004 to 2012. They showed that the categorization system became stable over time, especially with the introduction of top-level classification elements. Building on this work, Bairi, Carman and Ramakrishnan (2015) also looked at Wikipedia's evolution; they showed that much of the evolution of Wikipedia's knowledge organization was about maintaining overall knowledge coherence rather than adding new knowledge. For example, there has been a 25% increase in the number of categories over the 2012-2014 period compared to a 12% increase in the number of articles. Likewise, the number of disambiguation pages has increased by 13%. Both of these analyses point to change coming in the form not just of additional concepts or categories but change coming from ongoing maintenance that is much more active in bottom-up-derived KOSs. This source of change is termed gardening.

While Wikipedia's knowledge organization does mirror that of more professionally curated systems (Suchecki et al. 2012), it is not captured with explicit semantics. That is, there is no official version defined using a formal representation language (e.g., SKOS or OWL). This is, however, beginning to change with the introduction of Wikidata (Vrandečić and Krötzsch 2014). Wikidata provides a structured data version of much of the information available within Wikipedia's infoboxes. The information is present in a standard instance and class hierarchy mirroring RDF(S). However, with more formal semantics, applying them consistently becomes challenging with over 16,000 active contributors (https://www.wikidata.org/w/index.php?title=Wikidata:Statistics&oldid=320545760). This is documented in the discussion page about help for membership properties (https://www.wikidata.org/w/index.php?title=Help_talk:Basic_membership_properties&oldid=260792038) where Wikipedians discuss how to present help for how to interpret properties like "instance of," "sub class of," and "part of." Likewise, these more formal properties may be slower to become available across the totality of the knowledge organization. For example, property constraints are just currently being developed and are applied on only a small subset of Wikidata entity descriptions (Erxleben et al. 2014).

Wikidata also sees a speed of change in its knowledge that is at least an order of magnitude more than most traditional KOSs. Since its inception, it has had more than three hundred fifty million edits (https://www.wikidata.org/w/index.php?title=Wikidata:Statistics&oldid=320545760); this stems from both the number of contributors as well as major usage of automated agents. It is important to note that Wikidata like Wikipedia is systematically versioned.

Wikidata points at two new sources of change: 1) incremental and high speed modifications rather than sequential releases; and 2) progressive formalization rather than consistent and well known formalization (one is not guaranteed that formal semantics is applied throughout and for all concepts).

While we have focused on Wikipedia and Wikidata as exemplars of the crowd construction of KOSs, this happens in other sites such as LibraryThing, a website for books (Heymann, Paepcke and Garcia-Molina 2010). Even if the construction of a KOS is not organized by the crowd, it is increasingly likely that the crowd will be part of its construction as was discussed in the Association for Information Science and Technology 2016 panel "Crowdsourcing Approaches for Knowledge Organization Systems: Crowd Collaboration or Crowd Work?"

3.2 Automated knowledge base construction

As noted earlier, software and in particular automated extraction systems are now an important part of creating KOSs. This is shown by the extensive use of bots in Wikidata (https://www.wikidata.org/wiki/Wikidata:Bots). Additionally, there are a wide variety of systems that automatically construct knowledge bases by both automated text extraction and data integration (Suchanek et al. 2014). It is worth noting that these systems acquire both terminologies—the t-box—as well as statements conforming to the acquired knowledge organization—the a-box. There is a long history of automated knowledge classification, extraction, and markup (Gangemi 2013).

However, this is an increasingly active field because of application of knowledge bases in large-scale search by the likes of Google and Microsoft under the heading of knowledge graphs (Dong, et al. 2014). According to Nickel et al. (2016), these knowledge bases can contain billions of facts and thousands of types. Automated knowledge base construction employs a variety of techniques from open information extraction to link prediction and data integration. Indeed, Biega, Kuzey and Suchanek (2013) detail the usage of over thirty different extraction algorithms by the YAGO system. Moreover, these systems not only apply multiple algorithms but also use multiple sources that in turn are built of subsequent sources (Groth 2013). For example, the NELL system derives its knowledge organization by crawling millions of web pages constantly (Mitchell et al. 2015).

These automated mechanisms for constructing KOSs introduce three important sources of change. The first source is that changes to algorithms can impact the resulting knowledge organization. Even if the source of data were to remain constant if a particular pipeline extractor is changed, it can impact the results. This is not too dissimilar to the impact a change in knowledge engineer can have but on a larger scale and in a more opaque and diffuse fashion. The second source of change is the breadth of underlying data and it is (sometimes) unclear provenance that can be used to build a KOS. No longer is a KOS derived from multiple sourced scientific articles as in the case of MESH or GO but instead from the Web as whole or multiple independent sources. Furthermore, a KOS can be designed to constantly update itself as new sources become available. An important variant of this later source of change is the impact that underlying data can have when used to train algorithms, which are in turn used to build and maintain a KOS. Thus, as the source of training data changes so does the KOS.

To summarize, we see the following new sources of change:

1. Gardening (i.e., ongoing maintenance);
2. Incremental contributorship;
3. Progressive formalization;
4. Software and automation;
5. Integration of large numbers of data sources; and,
6. Variance in algorithm training data.

4.0 Discussion

Our review of sources of change to KOSs and how people contribute change has revealed a few interesting patterns. Some of the older models like *DDC* and *LCSH* were broad in scope and had wider distribution. Along with that, they approached change with a relatively conservative approach. This seems logical as the changes they were making affected many stakeholders and rolling back (restoring) changes would cause problems to downstream consumers. Also, as noted, many of the changes were regarding sensitive topics about people and culture.

As new KOSs are developed, new significant sources of change to models are emerging: non-professional contributors and software. The positive aspects to harnessing these new sources is their volume and efficiency. There may also be a perception that these processes remove much of the human element from KOS design and the output is somehow neutral, unbiased, and accurate. The fact is that bias will continue to permeate through these processes. As an example, the uneven distribution of topics in Wikipedia could lead to bias appearing in a KOS that used it as a source. As far as accuracy goes, algorithms are dependent on quality training data and the best practices for developing training data is to have a quality selection workflow with heavy involvement from subject matter experts. This could ultimately result in another form of bias.

The KOSs that result from crowdsourcing and software will not be absent human supervision. However, the roles that humans play will be quite different from before. As described earlier, the traditional KOS followed a production model similar to an editorial board with section specialists and subject matter experts. There will still be roles for specialists and experts in new models. The ideal will likely be hybrid models where software does much of the heavy lifting to detect and recommend new concepts. In these workflows, specialists will need to verify recommendations and make sure that they are in the scope of the model and the applications that are using it. Experts will also be needed to contribute to some linguistic aspects to the concepts. For example, experts will be needed to select preferred terms and synonyms.

If modern and future KOS development is going to involve a hybrid approach, it means that there are additional design considerations. An expert reviewing a suggestion for a change will want to know the source and provenance of the suggestion. This would likely start with categories of recommendations. Did the suggested concept come from an algorithm, a crowdsourced model, or a user suggestion? And from there, they will want to capture the specific source and record provenance information about the particular recommendation that will be saved with the metadata for that concept. Additional information will have to be carried along for suggestions that come from trained algorithms. Presumably, for an expert to receive a notification that a new concept needs to be reviewed, that recommendation would have had to have passed a certain 'likelihood' threshold. The editor would want the ability to review the confidence score for that concept and put it into context with other similarly suggested terms.

5.0 Conclusion

We have cataloged nine sources of change that impact the construction and design of knowledge organization systems: changing norms, political influence, new developments, maintenance; incremental contributorship; progressive formalization; software and automation; integration of large numbers of data sources; and variance in algorithm training data. Six of these changes are largely the result of new mechanisms for KOS construction, in particular, the introduction of crowdsourcing and automation. We hope that this list helps those responsible for designing, building, and maintaining KOSs and reflects on appropriate policies, guidelines, and development practices to deal with change.

Note

1. At the time of this writing in July 2016, the Library of Congress plans to stop using the term "illegal alien," as they are also receiving public comment on the matter and the passage of the related Legislative Branch Appropriations Bill is pending.

References

Aguilera, Jasmine. 2016. "Another Word for 'Illegal Alien' at the Library of Congress: Contentious." *New York Times* July 22, 2016. http://www.nytimes.com/

Bairi, Ramakrishna B., Mark Carman and Ganesh Ramakrishnan. 2015. "On the Evolution of Wikipedia: Dynamics of Categories and Articles." In *Ninth International AAAI Conference on Web and Social Media*.

Berman, Sanford. 1971/1993. *Prejudices and Antipathies: A Tract on the LC Subject Heads Concerning People*. Metuchen, NJ: Scarecrow. http://www.sanfordberman.org/prejant.htm

Biega, Joanna, Erdal Kuzey and Fabian M. Suchanek. 2013. "Inside YAGO2s: A Transparent Information Extraction Architecture." In *Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, May 13-17, 2013*, eds. Daniel Schwabe, Virgílio Almeida and Hartmut Glaser. New York: ACM, 325-8.

Bushman, Barbara. 2016. "Re: Questions about MeSH/RDF," email correspondence with authors July 21, 2016.

Bushman, Barbara, David Anderson and Gang Fu. 2015. "Transforming the Medical Subject Headings into Linked Data: Creating the Authorized Version of MeSH in RDF." *Journal of Library Metadata* 15, nos. 3-4:157-76

Committee Report, Legislative Branch Appropriations Bill, 2017. 114th Congress, United States https://www.congress.gov/congressional-report/114th-congress/house-report/594/1.

Dong, Xin, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun and Wei Zhang. 2014. "Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion." In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, August 24-27, 2014*. eds. Sofus Macskassy and Claudia Perlich. New York: ACM, 601-10.

Elsevier. 2016a. "Embase Content." https://www.elsevier.com/solutions/embase-biomedical-research/embase-coverage-and-content

Elsevier. 2016b. "Emtree Terms Added and Changed (January 2016)." Unpublished document.

Erxleben, Fredo, Michael Günther, Markus Krötzsch, Julian Mendez and Denny Vrandečić. 2014. "Introducing Wikidata to the Linked Data Web." In *International Semantic Web Conference, Riva del Garda, Italy, 19-23 October 2014,* ed. Peter Mika, Tania Tudorache, Abraham Bernstein, Chris Welty, Craig Knoblock, Denny Vrandečić, Paul Groth, Natasha Noy, Krzysztof Janowicz and Carole Goble. New York: Springer, 50-65.

Gangemi, Aldo. 2013. "A Comparison of Knowledge Extraction Tools for the Semantic Web." In *The Semantic Web: Research And Applications: 7th Extended Semantic Web Conference, ESWC 2010, Heraklion, Crete, Greece, May 30 - June 3, 2010*. Berlin; New York: Springer, 351-66.

Green, Rebecca. 2015. "Indigenous Peoples in the US, Sovereign Nations, and the DDC." *Knowledge Organization* 42: 211-21.

Groth, Paul. 2013. "The Knowledge-Remixing Bottleneck." *IEEE Intelligent Systems* 28, no. 5: 44–8. doi:10.1109/MIS.2013.138

Heymann, Paul, Andreas Paepcke and Hector Garcia-Molina. 2010. "*Tagging Human Knowledge*." In *Proceeding of Third ACM International Conference on Web Search and Data Mining (WSDM 2010), New York City, 3-6 February 2010*, eds. Brian D. Davison and Torsten Suel. New York: ACM, 51-60.

Klein, Michiel. 2004. "Change Management for Distributed Ontologies." PhD thesis, The Netherlands Research School for Information and Knowledge Systems.

Knowlton, Steven A. 2005. "Three Decades Since Prejudices and Antipathies: A Study of Changes in the Library of Congress Subject Headings." *Cataloging & Classification Quarterly* 40, no. 2:123-45.

Library of Congress. "SACO Funnels." https://www.loc.gov/aba/pcc/saco/funnels.html

Mitchell, Tom M., William Cohen, Estevam Hruschka, Partha Talukdar, Justin Betteridge, Andrew Carlson, Bhavana Dalvi Mishra, et al. 2015. "Never-Ending Learning." *AAAI Conference on Artificial Intelligence*: 2302-10.

Nickel, Maximilian, Kevin Murphy, Volker Tresp and Evgeniy Gabrilovich. 2016. "A Review of Relational Machine Learning for Knowledge Graphs." *Proceedings of the IEEE* 104, no. 1: 11-33.

Qin, Sonia. 2016. "Library of Congress to Replace Term 'Illegal Aliens.'" *The Dartmouth*. http://www.thedartmouth.com/article/2016/03/library-of-congress-to-replace-term-illegal-aliens/

Suchanek, Fabian, James Fan, Raphael Hoffmann, Sebastian Riedel and Partha Pratim Talukdar. 2013. "Advances in Automated Knowledge Base Construction." *SIGMOD Records Journal*.

Suchecki, Krzysztof, Alkim Almila Akdag Salah, Cheng Gao and Andrea Scharnhorst. 2012. "Evolution of Wikipedia's Category Structure." *Advances in Complex Systems* 15, no. supp01: 1250068.

U.S. National Library of Medicine. 2015a. "Fact Sheet, Medical Subject Headings (MeSH)." https://www.nlm.nih.gov/pubs/factsheets/mesh.html.

U.S. National Library of Medicine. 2015b. "MeSH Vocabulary Changes." https://www.nlm.nih.gov/mesh/intro_voc_change.html.

Voss, Jakob. 2006. "Collaborative Thesaurus Tagging the Wikipedia Way." arXiv preprint cs/0604036.

Vrandečić, Denny and Markus Krötzsch. 2014. "Wikidata: A Free Collaborative Knowledgebase." *Communications of the ACM* 57, no. 10:78-85.