Knowledge organization trends in library and information studies: a preliminary comparison of the pre- and post-web eras

Kristie Saumure and Ali Shiri
School of Library and Information Studies, University of Alberta, Canada

Abstract.
Qualitative analyses were used to launch a preliminary exploration of the dominant knowledge organization (KO) trends in the pre- and post-web eras. Data for this study was assembled by searching the Library, Information Science, and Technology Abstracts database for articles that have used the term ‘knowledge organization’ or ‘information organization’ in their titles, abstracts, or descriptors. Taken as a whole, these preliminary results suggest that the content of the KO literature has shifted since the advent of the web. Although classic KO principles remain prominent throughout both eras, the presence of new content areas, such as metadata, denotes a shift in KO trends. In the pre-web era, the literature was related in large part to indexing and abstracting. In contrast, cataloging and classification issues dominate the landscape in the post-web era. The findings from this paper will be of particular use to those interested in learning about upcoming trends in the KO literature.

Keywords: information organization; knowledge organization; world wide web

1. Introduction

Since Sir Anthony Panizzi’s mid-nineteenth century development of a plan for organizing materials within the British Library, knowledge organization (KO) has been considered a core principle within library and information studies (LIS) [1]. However, recently and particularly with the rapid development of the internet, new trends have emerged in this area of study. By reviewing and analyzing relevant, pre-selected articles in the Library, Information Science, and Technology Abstracts (LISTA) database, this research will provide a preliminary exploration of the dominant trends in
KO that have emerged in the last 40 years – paying particular attention to the differences between
the pre- and post-web eras.

This particular study builds on the work of both Hope Olson [2] and I.C. McIlwaine [3]. In her
longitudinal examination of how the information organization literature was represented in Library
Quarterly, Olson [2] found that this particular journal focused on issues pertaining to cataloging
codes, the cost of cataloging, and subject access, along with card catalog and microform technolo-
gies (though with minimal focus on computer technologies). It is expected that the current study
will differentiate itself from Olson’s [2] excellent work by casting a broader net. By exploring knowl-
edge organization issues in a database (which indexes multiple journals), as opposed to a single
journal, this study has the potential to find a greater variety of KO trends. In particular, by examin-
ing the technologically focused LISTA database, there is an increased likelihood that KO articles
related to computers will be found.

In contrast to Olson’s [2] study, McIlwaine’s [3] important work covers a shorter timeframe
(1998–2003; but see also McIlwaine and Williamson 1999 for a review of trends from 1988–98). This
work highlights what she views as the recent trends in KO: universal classification systems, inter-
operability, elimination of bias in knowledge organization, the internet/search engines, resource dis-
covery, terminology, thesauri, visual presentation, and the commercial environment. The current
study takes a different approach than that of McIlwaine [3] by looking at KO trends over a longer
time period (1966 to the present), as well as examining trends that have emerged since McIlwaine’s
study was published in 2003. It is also important to point out that this study is completely different
from the two aforementioned articles in its systematic comparison of pre- and post-web knowledge
organization trends. As a result, this study has the potential to provide insight into how KO trends
have shifted since the advent of the web.

2. Research questions

Consistent with the goal of launching a preliminary investigation of how the KO literature has
developed over the past 40 years, the following research questions are proposed.

What trends have emerged in the knowledge organization literature over the past 40 years?

a. What were the dominant trends in the pre-web era?

b. What were the dominant trends in the post-web era?

3. Methodology

3.1. Data collection

Data for this study was assembled by searching LISTA (formerly Information Science and Technology
Abstracts) for articles that have used the term ‘knowledge organization’ or ‘information organization’
in their titles, abstracts, or descriptors. Although ‘knowledge organization’ and ‘information organi-
zation’ are limited in scope and will not be inclusive of all KO trends (for example, some cataloging
articles may not have used either of the two aforementioned terms), we felt that it was important to
view the trends in the context of authors who have used the terms ‘knowledge organization’ and/or
‘information organization’ to describe their work. This would provide us with a window into author-
described KO trends (that is, articles that the authors themselves felt were about KO). At this point,
it should be acknowledged that the term ‘knowledge organization’ has been and is used very broadly
and sometimes vaguely by a variety of communities including LIS, computer science, cognitive psy-
chology and philosophy. There are researchers and writers outside the LIS domain that use this term
with no prior exposure to the long standing tradition of thesauri and classification schemes which
have been around for decades. For example, computer science writers, and in particular artificial
intelligence experts, do use the term knowledge organization and representation but with different connotations and meanings. In addition, there are various knowledge organization practices that are not necessarily based on the use of such knowledge organization systems as thesauri, classification schemes, subject heading lists and taxonomies. For instance, following the advent of the web, many web-based search services have been using the term knowledge organization to refer to techniques such as automatic categorization, auto-classification, clustering, automatic text summarization, and data and text mining.

The creation date of this database (i.e., 1966) was used as the starting point for the content analysis [4]. This database provides a manageable starting point for conducting a content analysis of the knowledge organization literature and is one of the longest standing databases of information science materials. LISTA includes pieces related to ‘librarianship, classification, cataloging, bibliometrics, online information retrieval, information management, search engines, printed and electronic information sources, the information industry, scholarly communication, and electronic publishing’ [4]. For the purpose of this LIS-focused paper, knowledge organization will be defined as those areas of study related to cataloging, classification, indexing, abstracting, thesauri, and authority lists [5]. Initially, 320 articles were located using the search terms described above – 230 articles using the term ‘information organization’ and 90 articles with the term ‘knowledge organization’. However, a detailed review of the articles revealed that there were numerous false hits amongst these articles. Hence, of the 320 initially selected articles, 101 were deemed inappropriate for the purpose of this analysis. Articles were removed from consideration in the content analysis because:

- they were non-refereed articles (i.e., book/website/article reviews, conference announcements, biographies of famous librarians, and awards announcements);
- the article was not really about knowledge organization, but was instead a more macro piece of which knowledge organization was only a peripheral part (e.g., knowledge organization as one educational aspect within library and information studies);
- the article was not about knowledge or information organization as a process, but was instead about knowledge or information organizations as entities (e.g., Bell Canada);
- and finally, due to language constraints on the part of the researcher, articles were removed from consideration if they were not in English.

For the purpose of this study, the post-web era was marked by the advent of the first internet browser (Mosaic) to run on a Windows platform in 1993 [6]. Hence, articles retrieved were classified as pre- or post-web, depending upon whether they were published before or after the advent of Mosaic. There were 31 pre-web articles and 188 post-web articles included in the analyses.

3.2. Procedural framework: thematic analysis procedure

To provide a deeper picture of the KO trends, a qualitative content analysis was also completed. Here, a thematic analysis of the KO trends noted in the titles and abstracts of these articles was completed; where necessary, the full article was also consulted to ascertain the types of trends identified in a given article. Unlike Järvelin and Vakkari [7], who use a pre-existing classification system to conduct their content analyses, this content analysis was influenced by grounded theorists [8] in that there was no attempt to establish a framework for the trends ahead of time. Instead, the researcher waited for the trends to emerge from the data. The trends were then examined in the context of how they differed before and after the advent of Mosaic.

Results of the literature search were imported into Microsoft Word to facilitate the qualitative analysis of the trends that have emerged in the KO literature since 1966 – paying particular attention to the differences that exist in the pre- and post-web eras. Identified trends were compiled in a Microsoft Excel spreadsheet to provide an overall picture of the dominance of particular KO trends.
4. Results and discussions: thematic analysis

The following broad-level trends were identified in the qualitative analysis of the targeted KO literature:

- Organizing corporate or business information
- Machine-assisted knowledge organization
- Librarians as knowledge organizers of the web
- Interoperability
- Cataloging and classification
- Classifying web information
- Digital preservation and digital libraries
- Metadata applications and uses
- Cognition
- Education
- Indexing and abstracting
- Thesauri initiatives

The ensuing discussion will provide insight into how these trends have evolved since the post-web era – and will point to sub-trends within these broader trends. For a representation of the prominence of these themes in the two eras, see Figures 1 and 2 (note: terms not appearing in the pie chart were not represented in that particular era).

4.1. Organizing corporate or business information

Discussions of how to organize information for corporate or business interests were not common in the reviewed pre-web KO literature, but this has changed in the post-web era. Recent research has
noted that some organizations have begun to build taxonomies in order to provide more efficient access to web-based business knowledge [9], while others have looked towards the creation of ontologies to help them manage corporate events – bringing together and organizing events from multiple sources [10]. Increasingly, the plethora of emails received by employees has also been seen as an issue that corporations must address. There is awareness that this information must be classified in order to ensure efficient retrieval – though one study points out that individuals prefer to categorize their own emails in order to ensure that they are able to find pertinent information [11]. Another study looked at the bulk of emails with respect to archiving – corporations are increasingly wasting virtual space with irrelevant emails (i.e. not related to corporate interests). It has been pointed out that computer programs have the capacity to sort relevant from irrelevant information, allowing the corporation to choose not to archive those emails that are not of immediate interest to the business or corporate body [12]. Finally, the need to evaluate the organization of corporate information is also becoming more common [13]. More specifically, some research has noted that organizing records in a decentralized environment is inefficient (often lacking specific standards across departments) and can result in records being more difficult to retrieve [14].

4.2. Machine-assisted knowledge organization

During the pre-web era, steps were being taken towards the use of machines in organizing knowledge. This was perhaps most evident in the indexing and abstracting field. Creation of programs, such as FASIT (Fully Automatic Syntactically Based Indexing of Text), revealed that the ability of computers to index materials was being explored [15]. In other areas of the indexing and abstracting field, the capacity of computers to build abstracts from the full text of a resource was being pursued [16]. Evaluation of the computer’s capacity to effectively organize information was also a common theme in the knowledge organization literature, with some pointing out that computers were as effective at indexing as humans [17].
With the advent of the web, the amount of information to which people were exposed exploded and the location of requisite information became increasingly difficult. As a result, it became increasingly important to explore the computer’s capacity to use meaning and context when classifying materials [18–20]. Projects, such as TDT (Topic Detection and Tracking) illustrated the ability of computers to filter news stories, so that they were categorized into the appropriate topic [21]. Other projects looked at the capacity of computers to index according to genre – thereby providing greater context for the resource [22]. The ability of computers to map keywords found in the text of a document to the correct term in a thesaurus was also reviewed [23–25]. The National Library of Medicine’s MetaMap System, for example, was able to map terms from the free text of a document to their Unified Medical Language System (UMLS) metathesaurus [23]. In addition, as the use of metadata became increasingly pervasive, means for automatically attaching and harvesting metadata were explored [26, 27]. Using computers to cluster semantically similar concepts was also explored for such purposes as creating thesauri or classifying materials in a digital library [24, 28, 29]. In an evaluative capacity, there was a critique by some that computers did not index materials in ways that were always semantically meaningful to searchers [30, 31].

4.3. Librarians as the knowledge organizers of the web

Despite some early fears, librarians in the post-web era are considered to be an essential service – in large part due to their knowledge organization skills [32]. Indeed, some suggest that as digital libraries and the digital world generally grow in complexity, human classification will become increasingly important [33, 34]. This is perhaps a function of the fact that computers will never truly be able to capture the subtle semantic nuances that are present in language.

4.4. Interoperability

During the pre-web era the term interoperability does not occur in the records analyzed. However, after the development of the web, the use of the term interoperability became pervasive in the reviewed literature (see concluding remarks for discussion of interoperability issues that prevailed in pre-web era, but were not analyzed in this particular project). Increasingly, organizations wanted to share their digital resources with other organizations [35]; this, however, required the development of metadata standards, schemas, and crosswalks [26, 36–43]. METS (Metadata Encoding and Transmission Standard), Dublin Core, and XML protocols were all considered particularly important to facilitating the sharing of information across repositories (see [36, 38, 44] for more details). METS works as a digital archiving tool which enables the movement of digital objects between digital repositories, but also across the web [36]. Dublin Core was considered important to the interoperability movement because it essentially created a common metadata structure for describing a document – hence enabling this common metadata structure to be shared amongst repositories or digital libraries [38]. Finally, XML protocols are considered essential to facilitating metasearches across repositories [44]. In addition to enabling users to search across multiple repositories, the creation of ‘interoperable’ metadata enabled robots to harvest metadata from a range of repositories; interest in this capability was spawned after the creation of the Open Archives Initiative Protocol for Metadata Harvesting (OAI/PMH; 42). The relative recentness of all these articles (i.e. 2005 and 2006) suggests that this issue is only now coming to the forefront of the knowledge organization field.

4.5. Cataloging and classification

Not surprisingly, cataloging and classification were common areas for discussion in the KO literature. Interestingly, it was topics related to classification that appeared to be more common during the pre-web era. For example, there were discussions related to the issue of mutual exclusivity when classifying material [45], while others examined the use of cosine and single matching functions for classifying information [46].
During the post-web era, articles connected to cataloging and/or classification were quite common. Related to cataloging, there were discussions associated with the cataloging of electronic materials. For example, one article explored the expected knowledge of catalogers with respect to their ability to catalog electronic materials [47]. Still other articles reflected upon the use of the web to provide tools for cataloging. An evaluative study at the University of Colorado-Boulder looked at whether catalogers were using the Library of Congress online resource ‘Classification Web’ to aid in their cataloging processes. Also explored was the issue of how Electronic Resource Management Systems (ERMs) would influence serials cataloging in the future. Not all discussions of cataloging during the post-web era were specifically related to the web. For example, some reviewed procedures related to copy cataloging, some examined the use of international MARC records, others looked at the need to understand manuscript cataloging, and still others discussed how to catalog the increasingly prevalent graphic novel [48–51]. Specific discussions of new issues relating to AACR2 were also present in the post-web knowledge organization literature [52–54].

Relating more specifically to classification were critiques of some current schemes, as well as future directions for resource classification. One particularly critical discussion was aimed at the United Nations Classification Scheme – the authors suggested that decentralization in the scheme’s development and maintenance was related to flaws in its simplicity, use of mnemonics, brevity, and collocation [55]. In terms of future directions, many authors came forward with ideas. There were discussions related to using vertical integration to classify scientific knowledge, using classic classification principles (e.g. relevance, precision) to classify research information, using patent co-citations to classify patents, using seminal mnemonics to create taxonomies, looking beyond DDC (Dewey Decimal Classification) and Library of Congress Classification to create new classification structures, and classifying undergraduate information by course and not subject [56–61]. The importance of context when classifying information is found in a discussion of classifying by genre [62]. With respect to the digital or post-web era in particular, recommendations were made for adopting techniques to classify digital resources for the classification of artifacts. Related to this classification of digital resources, users of Walden’s Path point to the utility of bringing together disparate links into a single path [63]. Also increasingly common in the area of classification is the use of topic maps [64–66]. In one particularly interesting example, literary associations with the prominent literary figure Thomas Mann are provided by means of a topic map [65].

4.6. Classifying web information

In this post-web era, the classification of web information warrants a discussion separate from that of more general cataloging and classification issues due to its pervasiveness in the literature. Increasingly, metadata is becoming a key tool for organizing and providing access to web-based knowledge [67–69]. This discussion of metadata takes place in a number of ways. One issue of interest relates to the conversion of print-based bibliographic records into a format more befitting the classification of information on the web, such as Dublin Core [70]. Importantly, metadata schemas also become a powerful tool for retrieving web based information [44, 71] – problematically however, there is some indication that metadata schema and metatags are not well used and hence access to information is reduced [72].

In addition to metadata, there is also a sense that faceted classification is becoming a key means of organizing information on the web [73, 74]. Faceted classification has taken on a particularly prominent role with respect to organizing product information on commercial websites [75]. Furthermore, faceted classification can also be considered a useful tool for organizing information that has been gathered in the data mining process (e.g. linking terms found in data mining to terms found in a thesaurus [76]). In addition to faceted classification, there has also been a move towards embracing library-specific organizational techniques – perhaps with the help of a librarian [77]. Dewey Decimal Classification, for example, is considered to be of value when browsing large collections [78]. Ontologies are considered valuable to classifying web information in that they aid in enhancing interoperability – bringing together resources from multiple sources [10, 79, 80].
Taxonomies can be considered another tool useful for organizing web-based information. For example, taxonomies provide an excellent means for organizing subject-specific information into an easily navigable format [81]. The utility of taxonomies for displaying web information can also be found when examining Yahoo’s site – which uses a taxonomy/subject hierarchy to classify its indexed information [82]. Moreover, the Integrated Museum and Archives System Taxonomy is also considered to be of significant utility when searching for cultural resources across museums and archives [83]. Finally, efforts to create a taxonomy of computer science terms using Dewey Decimal Classification and the IEEE Web Thesaurus have also proven to be successful [84].

4.7. Digital preservation and digital libraries

Issues related to the digital preservation of materials were not apparent in the pre-web era, but came to the forefront of knowledge organization discussions in the post-web era. In particular, it became increasingly apparent that the web was an excellent means for making tangible resources available virtually across geographically or topically disparate areas [35, 43, 85]. As this utility became apparent, it was perhaps inevitable that suggestions for enhancing access to these digital resources came to the fore. For example, clustering techniques were advanced as a tool for improving the navigational structure within a digital library [86]. Others pointed out the need for standards in creating and formatting metadata [36]. Yet another suggestion related to how information should be organized within a digital library, suggesting that a task-based approach may be more efficient than either subject or alphabetical approaches [87]. As digital collections became increasingly common, evaluations of their utility and content began to emerge. While one study used a user-centered analysis to evaluate how well a digital library was organized [88], another study examined the core trends related to digital libraries (one of which was knowledge organization systems) [89].

Perhaps the most common discussion area around knowledge organization and digital collections relates to metadata. There is a belief that metadata is an essential component for accessing information in digital collections [90]. Interest in types of knowledge organization systems used to help generate metadata is also common – with a Canadian study suggesting that thesauri, lists of subject headings, and classification schemes are the most frequently used [91]. Related to the earlier described concept of interoperability, metadata are also considered to be a key tool for ensuring that metadata from one digital collection can be used across multiple instances [9]. Indeed, METS is proving to be one way of ensuring this interoperability [92].

4.8. Metadata applications and uses

Not tied specifically to instances of digital collections, metadata applications are proving to be a topic of discussion all on their own. SKOS (Simple Knowledge Organization System) Core, for example, is considered to be an important tool for working with the semantic web – enabling translation from more ‘traditional’ classification systems to a semantic web language [35]. The Research Support Libraries Programme Collection-Level Description Schema enables the identification of different attributes within a library collection [93]. Mentioned earlier in relation to the discussion of interoperability, METS has become the standard by which metadata is encoded with respect to digital library objects [92]. Also important, Dublin Core has become an important tool for tagging a wide range of web resources [37, 39, 94]. Encoded Archival Description (EAD) has been used to create metadata for the purpose of improving archival control [95]. The creation of a metadata schema registry (InterPARSE2) has also proved important to the archiving community – providing a locale where archivists can learn to work with pre-existing metadata schemas or develop their own [56, 96]. Others, recognizing the utility of metadata for tracking resources, have developed tools (e.g. METEOR) that enable metadata gathering or harvesting [97]. Finally, evaluative measures are also beginning to emerge with the development of MODAL (Metadata Objectives and principles, Domains, and Architectural Layout) – which allows users to look at a metadata scheme's functions [98].

Also present in the literature are discussions of how metadata has been used for different applications. For example, there are reports of metadata being used to enable the sorting of key document
characteristics, so that users are able to look for information in a manner that best suits them [99]. Others have used metadata to manage the digital objects present in a multimedia tour of Loch Lomond and the Trossachs National Park [100]. In the e-commerce industry, metadata has been created that enables the user to learn about the quality of a resource. Others have created metadata to help organize their paper documents [101]. Essentially, in these instances, metadata is being used to meet the goal of providing optimal access to resources [102].

4.9. Cognition

Knowledge is organized so that human beings are able to find it in a more expedient manner. With that in mind, it is perhaps not surprising that models of human cognition have been important muses for creating knowledge organization systems [103]. Use of cognitive models has remained important to designing knowledge organization systems in both the pre- and post-web eras. For example, in the pre-web era, a study of how people group information (on a cognitive level) becomes a tool for assessing how well a batch retrieval system works [104]. Other studies in this era looked at how cognitive differences influence the way individuals assess the relevance of a document and how one’s cognitions may influence whether you prefer access points other than title, author, and subject [105, 106]. After the advent of the web, the use of cognitive models looked quite similar in nature. One study examined how children classify information in order to determine how information should be organized in a children’s library, while another study looked at how belief changes may influence how people cognitively organize information [107, 108]. Despite the continuing use of cognitive models, there has been some critique that these cognitive models do not account for the sociocultural differences which may influence the words people use to retrieve information [109].

4.10. Education

Another common trend in the knowledge organization literature is to assess the education being provided to knowledge organization specialists. Surprisingly, this trend seems to have taken place primarily in the post-web era (but see [110]). This might result from the increasing awareness from knowledge organization professionals that the web has in many ways changed the way that information is organized. For example, there is now a need to teach issues about how to ensure interoperability amongst digital libraries [68]. A greater focus on how to apply metadata is also considered to be a new competency amongst knowledge organization professionals [111, 112]. Other areas of focus when discussing knowledge organization include descriptions of knowledge organization courses in different geographical regions, as well as an understanding of what core competencies knowledge organization professionals need when entering the workforce [113–117]. There is also a sense that education related to knowledge organization needs to move outside of the librarian’s possession and into areas such as computer science [118].

4.11. Indexing and abstracting applications and procedures

Discussions of indexing and abstracting were quite common in the pre-web era. A great deal of attention was paid to ensuring the quality of the indexing. A study confirming which of two indexing procedures worked better was conducted in order to ensure the quality of retrieval [119]. Another study looked at the need to provide more terms when indexing an object as this would enhance the user’s chance of retrieval [120]. Devices, such as PRECIS (PREserved Context Index System) were also designed in order to help improve the subject indexing process at the British National Bibliography [121].

Post-web studies of indexing and abstracting are often related to the indexing of digital information. For example, one study looked at the use of Novel Space Partitioning (NSP) to index multidimensional non-ordered discrete data spaces, such as a genome database [122]. Another study discussed how the Laurin Thesaurus was used to index keywords and proper names, which aided in the indexing of digital newspaper clippings [123].
4.12. Thesauri initiatives

This theme points to the large scale use of thesauri following the advent of the web and the fact that thesauri were developed and used long before the web but usually by specific and individual information retrieval systems and not widely as is the case these days. One example of this attention was the creation of the Universal Source Thesaurus and how it might be used to create a concordance between classification schemes [124]. With the development of the web, the use of the thesaurus is coming to the forefront of knowledge organization studies. There are numerous examples of thesauri development in this era. The Chinese Thesaurus was created as a means of standardizing China’s retrieval language [125]. The National Library of Medicine created the MetaMap to enable the mapping of free web-based text to terms in the Unified Medical Language System Metathesaurus [23]. The Laurin Thesaurus was developed to enhance the indexing of digital clippings [123]. New trends in developing thesauri have also been emerging in the post-web era. For example, some attention has been paid to the development of thesauri using conceptual clustering methods [28]. Similarly, others have noted that bibliometrics might aid in the creation and maintenance of thesauri – using a semi-automatic approach [24]. Others have pointed out that terms found in the titles of documents might be used to glean terms that would be useful to include in the thesaurus [25]. Interest in thesauri has been revived to such an extent that review articles have now been published – examining the trends within web-based thesauri [126].

5. Concluding remarks

Taken as a whole, these results suggest that the content of the knowledge organization literature has shifted since the advent of the first graphical web browser. The qualitative analysis points to the significant role that metadata is now playing and is going to play within knowledge organization research and practice – something that points again to the evolving look of knowledge organization in the post-web era when compared with the pre-web era. Indeed, metadata as a concept runs through many of the core themes identified in the qualitative analysis. For example, within the ‘Machine Assisted Organization’ theme, attention is being paid to how computers can help generate, as well as harvest, metadata.

Within the domain of ‘interoperability’, it has been recognized that in the post-web era, metadata and accompanying metadata standards will be an essential component in sharing and harvesting information across repositories. It is worth pointing out that, despite the lack of analyzed papers that explicitly used the term interoperability, it cannot be said that this concept was not discussed or investigated prior to the advent of the web. The notion of interoperability, particularly subject interoperability, is not new. The discussion of issues related to the integration and mapping of thesauri, classification schemes and other types of controlled vocabulary has a long history. A wide range of projects have investigated subject interoperability using such terms in their titles as reconciliation of thesauri [127], integration [128,129], compatibility [130,131], merging [132], switching [133], and mapping [134,135]. Our analysis focused on the use of the term interoperability and its occurrence in the analyzed records.

Turning back to the discussion of metadata, it has also been acknowledged that metadata will play a key role in classifying the astronomical amount of information found on the web – making the information retrieval processes more manageable. In a related vein, metadata is a driving force in the creation of useable digital collections – enhancing both the browse and search features within these collections. Thesauri are helping to provide metadata indexers with the tools necessary to apply subject terms in a consistent fashion – again aiding in the retrieval process. It is perhaps the dominance of metadata in these knowledge organization domains that has led knowledge organization educators to point out the utility and necessity of teaching students about metadata.

This focus on metadata is not meant to suggest that pre-web knowledge organization principles are falling by the wayside. Indeed, discussions of AACR2 are still common. Moreover, pre-web knowledge organization methods are being used to create and apply metadata. Within the indexing
field, metadata is being used to index digital objects in a manner that bears resemblance to the indexing of print materials (e.g. using controlled vocabularies to index documents). As well, the earlier discussion related to metadata and thesauri also highlights how pre-web applications (i.e. thesauri) are being used to make metadata more useable. As the organization of knowledge and information continues to evolve in this post-web context, it seems evident that the relevance of core knowledge organization principles will remain, despite the shifting trends. These principles will most certainly help enhance both the browsability and searchability of emerging web-based environments, such as digital libraries, content management systems, institutional repositories, and virtual learning environments.

The findings from this paper will be of use to researchers, instructors, and students in library and information studies who are interested in learning about the direction that the KO literature is taking – offering particular insight into how things have changed since the advent of the internet. More specifically, researchers may find it useful for both generating new ideas and informing their literature reviews, while students may find it gives them insight into a new area of study. Furthermore, instructors who are developing knowledge organization courses may find the attention paid to defining new trends important to developing a ‘cutting edge’ knowledge organization course.

6. Limitations and areas of future research

Two key limitations should be pointed out with respect to this project. First, the analysis was only conducted using the results from a single search of a Library and Information Studies Database (i.e. LISTA). For that reason, articles that were not indexed in this particular database, but were for example indexed in Library and Information Science Abstracts (LISA), were not analyzed. For that reason, future studies should consider doing a comparable analysis using a different information studies database.

A second limitation of this research study relates to the terms used to search for articles that describe trends in the knowledge organization literature. Although the use of these terms was important in enabling us to examine the general discourse offered by authors who have used these terms as descriptors or in their titles or abstracts, there are certainly researchers outside the LIS domain who are using the terms in ways that fall outside the realm of longstanding knowledge classification systems, such as thesauri and subject heading lists. As a result, because this particular study only considered the broader terms ‘information organization’ and ‘knowledge organization’ when searching for pertinent articles, it might also be useful to consider a more narrowly focused study of terms, such as ‘classification’, ‘indexing’, and ‘thesauri’. This type of analysis has the potential to provide finer details regarding recent trends in knowledge organization that were not picked up in this broader analysis. Furthermore, utilization of narrower terms may result in more pre-web articles being recalled; hence, allowing for a more balanced comparison of the pre- and post-web literature.

Despite the limitations described above, we are confident that this project provides new insight into knowledge organization trends (pre- and post-web) and is a catalyst for the development of future research in this area. As expected, it differentiates itself from Olson’s [2] work by providing a more computer-focused examination of knowledge organization both today and in the past. Moreover, it updates the trends identified by McIlwaine [3] in her 2003 examination of current knowledge organization trends (e.g. using metadata to help organize information in digital repositories) while also pointing to how recent (i.e. post-web) trends are different than those found in the pre-web era.

References

Kristie Saumure and Ali Shiri


