Knowl. Org. 37(2010)No.1
J. H. Leong. The Convergence of Metadata and Bibliographic Control?
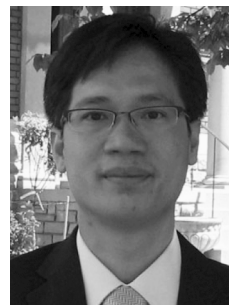
29

The Convergence of Metadata and Bibliographic Control? Trends and Patterns in Addressing the Current Issues and Challenges of Providing Subject Access

Jack Hang-tat Leong

University of Toronto Libraries, The Richard Charles Lee Canada-Hong Kong Library,
130 St. George Street, Toronto, Ontario, Canada, M5S 1A5, Canada,
<jack.leong@utoronto.ca>

Jack Leong is the Director of the Richard Charles Lee Canada-Hong Kong Library at the University of Toronto. Dr. Leong has degrees in computer science, education, comparative literature, and library and information science. Since 2006, he has been working as a librarian. His research interests include knowledge organization, subject access, Chinese Canadian studies, and science fiction. Dr. Leong serves on the Board of Directors of the Canadian Foundation for Asian Culture. He is also a standing committee member of the IFLA Section "Library Services to Multicultural Populations."

ABSTRACT: Resource description and discovery have been facilitated generally in two approaches, namely bibliographic control and metadata, which now may converge in response to current issues and challenges of providing subject access. Four categories of major issues and challenges in the provision of subject access to digital and non-digital resources are: 1) the advancement of new knowledge; 2) the fall of controlled vocabulary and the rise of natural language; 3) digitizing and networking the traditional catalogue systems; and 4) electronic publishing and the Internet. The creation of new knowledge and the debate about the use of natural language and controlled vocabulary as subject headings becomes even more intense in the digital and online environment. The third and fourth categories are conceived after the emergence of networked environments and the rapid expansion of electronic resources. Recognizing the convergence of metadata schemas and bibliographic control calls for adapting to the new environment by developing tools that exploit the strengths of both.

1.0 The trend of convergence

Resource description and discovery have been facilitated generally in two approaches, namely bibliographic control and metadata. Rooted in library and information science, standards of bibliographic control, such as the *Anglo-American Cataloguing Rules* (*AACR*) and the *International Standards for Bibliographic Descriptions* (*ISBD*), have mandated effective organization and arrangement of information, making it readily identifiable and retrievable. The practice and theory of this domain have emphasized the "systematic, uniform, and consistent approaches to describing intellectual or artistic content and physical characteristics" (Howarth 2005, 39). Metadata, a file management concept originated in the computer science community, has been conceived as an information management and retrieval tool to handle ever-increasing online information sources. The concept of metadata is often referred to as "structured data about data" deriving from its roots in computer systems. At the end of the twentieth century, the term was redefined by the library, archive, and museum information communities, warranting a "confluence in

30

Knowl. Org. 37(2010)No.1
J. H. Leong. The Convergence of Metadata and Bibliographic Control?

terminology and definition" between bibliographic control and metadata (Howarth 2005, 37). Discussing the convergence of these two approaches requires distinguishing the two concepts. In this paper, metadata retains its esoteric definition from the late 1990s, referring to structured data generated automatically for the description and recovery of resources stored in electronic systems.

The proliferation of digital publications creates challenges to both traditional bibliographic control and metadata schemas. As developed primarily for books and textual documents, traditional bibliographic control's capability to manage and provide access to networked electronic resources is questioned by the sheer number of electronic resources and other conditions, particularly the need for trained cataloguers (Chan 2005; Gross and Taylor 2005; Markey 2007). Nevertheless, the search results produced by information retrieval systems utilizing less structured metadata are often unsatisfactory. As a result, the metadata community has proposed various measures to improve performance of these systems. Among these measures, the most prominent ones seem to be building end-user thesauri, providing vocabulary mapping and adopting major subject headings and classification systems in subject metadata fields. These

strategies interestingly resemble a number of fundamental mechanisms in bibliographic control, suggesting a direction of cooperation between bibliographic control and metadata schemas in addressing the current challenges. Figure 1, which illustrates the emergence of common concepts and tools of bibliographic control and metadata schemas, demonstrates this converging trend visually. As a result of this convergence, subject access has been improved significantly with the synergy of human analysis and computer algorithms, controlled vocabulary and keywords, as well as users and trained cataloguers.

2.0 Challenges to knowledge organization

2.1. The advancement of new knowledge

Knowledge is being constructed on a continuous basis as new discoveries are made and new concepts are conceived. A brief review of the weekly list of new *Library of Congress Subject Headings* (*LCSH*) reflects this phenomenon. Even though it seems to be a subject heading system responding fairly conservatively towards new knowledge based on the practice of literary warrant, *LCSH* includes dozens, if not hundreds, of new subject headings every week (Cataloging Pol-
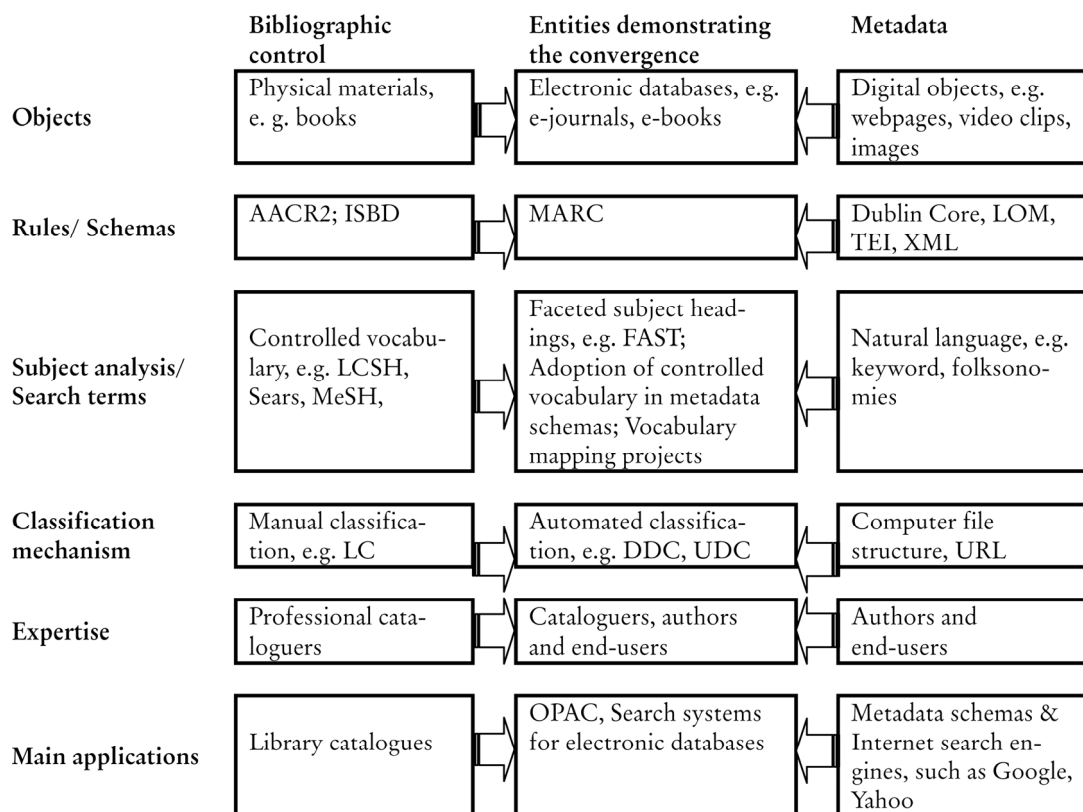
	Bibliographic control	Entities demonstrating the convergence	Metadata
Objects	Physical materials, e. g. books	Electronic databases, e.g. e-journals, e-books	Digital objects, e.g. webpages, video clips, images
Rules/ Schemas	AACR2; ISBD	MARC	Dublin Core, LOM, TEI, XML
Subject analysis/ Search terms	Controlled vocabulary, e.g. LCSH, Sears, MeSH,	Faceted subject headings, e.g. FAST; Adoption of controlled vocabulary in metadata schemas; Vocabulary mapping projects	Natural language, e.g. keyword, folksonomies
Classification mechanism	Manual classification, e.g. LC	Automated classification, e.g. DDC, UDC	Computer file structure, URL
Expertise	Professional cataloguers	Cataloguers, authors and end-users	Authors and end-users
Main applications	Library catalogues	OPAC, Search systems for electronic databases	Metadata schemas & Internet search engines, such as Google, Yahoo

Figure 1. The trend of convergence between bibliographic control and metadata schemas

icy & Support Office 2006). To complicate the issue, new knowledge is increasingly introduced in non-traditional formats. Besides books or textual documents of various kinds, new knowledge can be conveyed in a sound recording, a video clip, or an Internet website, to name only a few.

2.1.1 Organizing new knowledge

Beghtol (2004, 19) argues that classification systems can be termed "professional classifications" and "naïve classifications." While professional classifications, such as the *Dewey Decimal Classification* (*DDC*) and the Library of Congress *Classification* (LCC), attempt to organize and provide access to pre-existing knowledge, naïve classifications, usually developed by the domain-specific individuals who "have no particular interest in classificatory issues," primarily facilitate the discovery of new knowledge (Beghtol 2004, 19). Beghtol's argument suggests that new knowledge, before it becomes established and recognized in publications, is usually dealt with by naïve classifications within a particular context. This model of two-tier development of classification systems can explain interestingly the rise and popularity of numerous metadata standards for managing Internet resources. As "professional classifications" are not tailored for discovering and managing new information, which is often the content of Internet resources, the communities in various subject areas develop their own "naïve classifications" to deal with their knowledge management needs. Internet directories, search engines and folksonomies are thus developed for this purpose.

Professional and naïve classifications serve different contexts and purposes. Nevertheless, they depend on each other and have a "cyclical relationship" (Beghtol 2004, 22). Professional classifications provide one means of formulating research into new knowledge. As new knowledge is disseminated among scholars using naïve classification systems, it gradually attains the status of literary warrant and enters the professional classification systems for conception of new knowledge. Beghtol's model can illustrate the divide in terms of context and purposes of traditional bibliographic control (professional classifications) and less structured metadata applications (naïve classifications), such as folksonomies. Nonetheless, since the advancement of information and communications technology has been narrowing the time-lapse between new knowledge and knowledge harvested from literary warrant, there is a need to investigate and propose a new model for addressing the needs of managing

new knowledge. Because naïve classifications and professional classifications adopt similar structural principles for organizing knowledge, synergy between these two systems is possible and desirable. For example, as shown by the **subject analysis** and **classification** dimensions in Figure 1, from the metadata direction, many schemas have included data fields of controlled vocabularies and classification systems to improve the effectiveness of information search and retrieval (Lee-Smeltzer 2000). On the bibliographic control side, the move to a faceted syntax of *LCSH* and other subject headings and classifications demonstrates the employment of metadata techniques for facilitating the automation of the traditional cataloguing and classification processes. The convergence in the subject analysis and classification dimensions suggests unambiguously that both systems are modifying themselves while bringing them closer to each other.

2.1.2 An era of interdisciplinary knowledge

Besides the rapid creation of new knowledge, the interdisciplinary characteristics of new knowledge create challenges to standard information retrieval and classification systems. In the knowledge era with advanced information communication technology, people have access to a very wide spectrum of knowledge. Moreover, studies that cover several traditional disciplines contribute more and more to new knowledge. The issue of "interdisciplinarity" as a new information retrieval problem and its resolutions are discussed by McIlwaine (2000), who presents the revision of two interdisciplinary fields, namely tourism and environmental science, in the *Universal Decimal Classification* (*UDC*). Her discussion demonstrates the flexibility of *UDC*, a highly faceted classification system, in addressing changes and interdisciplinarity. In the online and networked environment, the ability to allow the co-existence of basic classes appears to be critical. Since faceted classification systems can enable synthesis, the networked information retrieval systems often engage the faceted approach if they require a classification system for providing an information organization schema that can benefit from computer manipulation.

The advancement of new knowledge and the increasing popularity of interdisciplinary studies have led to the rapid appearance of new subject terms in their common usage format. Since there is often a significant time lapse before a new concept is included in controlled vocabulary, it is necessary to examine the advantages and disadvantages of employing

32

Knowl. Org. 37(2010)No.1
J. H. Leong. The Convergence of Metadata and Bibliographic Control?

natural language and controlled vocabulary in information retrieval tools for the new environment.

2.2 The fall of controlled vocabulary and the rise of natural language

Librarians started to realize the significance of subject access to library materials in the end of the nineteenth century. Since then, the advantages and disadvantages of using natural language or controlled vocabulary for providing subject access have been debated continuously. The usefulness and strengths of controlled vocabulary for subject searching, particularly in terms of synonym and homonym control and term relationship, have been reinforced by various studies (Rowley 1994; Tillotson 1995; Gross and Taylor 2005). Most bibliographic control systems, therefore, continue to adopt this approach for providing subject access. However, the proliferation of networked electronic resources and the digitization of catalogue systems have enabled the natural language approach to play a more significant role in providing subject access to the electronic resources which are huge in number and fluid in structure. Indeed, the popularity of folksonomies and keyword searches enabled by database systems and Internet search engines asserts that natural language has become a dominant approach in some metadata applications that require less structure. This section briefly illustrates the tension between using natural language and controlled vocabulary in subject access and identifies the trend of adopting both approaches in new subject access systems.

2.2.1 The adaptations of natural language and controlled vocabulary in the online environment

As the expertise dimension of Figure 1 illustrates, a natural language approach often utilizes terms that are used by the authors or users to describe the primary content of the items for retrieval. Compared to the controlled vocabulary, the process of deriving natural language terms seems to be easier as people can use any words that they conceive. The assignment of subject headings to an item using controlled vocabulary requires trained cataloguers to examine a list of established headings and read through the instructions to find or construct the most appropriate headings. Moreover, in terms of access points, the natural language approach appears to be more comprehensive and exhaustive because users can use any, and as many, terms as they can conceive of for the item. Resource cataloguers thus have more choices to provide direct

access points that can capture the most common and recent terminology. Controlled vocabulary headings, limited by rules and the principle of literary warrant, appear to be less efficient in utilizing current terminology from the user's perspective. Controlled vocabulary systems do provide cross-references to refer from common terms not included in their lists to the authorized headings. However, not all natural and common terms are considered. Even if the common terms are included in the reference system, more steps are involved to access the item using these terms.

In the electronic and networked environment, the advantages of the natural language approach become very helpful to address the sheer number of resources available. Moreover, complemented by computer algorithms and statistical analysis, the performance of subject access systems using the natural language approach has been improved (Bates 1989). Nevertheless, the drawbacks of the natural language approach are noticeable in the new environment. First, the terms may not be consistent. Different people may use different terms to describe the same content. Moreover, resources on a single topic, such as food, can bear different terms as their subject entries, such as cooking, diet and grocery. Third, the natural language approach does not provide any cross-references among the subject terms. As a result, relationship and hierarchy of subjects cannot be determined. Finally, the choice of terms is limited because the authors and the users are familiar with a certain set of vocabulary only while a subject heading system usually provides cross-reference and index tools to enrich the list of possible subject entries. These drawbacks pose two challenges in subject access. Users may have to try a number of synonyms to identify an item that they know by subject. Moreover, when they wish "to retrieve all information, or as much information as possible, on a given subject, [they] must search all synonyms for that subject" (Chan 1994, 155).

In their discussion on folksonomies, a natural language approach of providing subject access, Guy and Tonkin (2006) observe similar drawbacks in what they call "sloppy tags" or "low-quality, redundant or nonsense metadata." To address these drawbacks, critics of folksonomies suggest some resolutions that exploit the strengths of subject analysis and access using controlled headings. Indeed, controlled vocabulary seems to be more consistent, uniform and unique. Moreover, the controlled vocabulary system is adopted by most libraries and therefore the catalogue systems using these headings can become standardized in subject access. Moreover, syndetic structure and subject author-

ity files have been established among the controlled headings, and they provide a rich set of vocabulary for the authors or users to describe a resource. Traditional catalogue systems using controlled vocabulary have been serving users to identify a known item or gather items on the same subject in a very efficient and seamless manner in the library environment for a long time. The strengths of controlled vocabulary, such as its higher performance in the measures of precision and recall, can be used to complement the drawbacks of the natural language approach in the development of new subject access tools that can embrace these two approaches simultaneously (Chan 2000). As demonstrated by the automation of *Dewey Decimal Classification* (*DDC*) and Faceted Application of Subject Terminology (FAST), controlled vocabulary systems are being modified to adapt to the online environment and to enable the engagement of the natural language and controlled vocabulary approaches. Koch (2000), for example, argues that quality-controlled subject gateways, which exploit both the controlled vocabulary and the deep-structured classification systems, provide quality measures, to support systematic resource discovery. The quality measures include using controlled vocabulary, together with a thesaurus, for indexing resources and deep-structured classification systems for providing advanced searching and browsing capabilities.

2.2.2 *Embrace the combination of natural language and controlled vocabulary*

Several tools and projects have been developed to facilitate the simultaneous engagement of natural language and controlled vocabulary. Buckland (1999), for instance, implemented the Unfamiliar Metadata Vocabularies Project that maps initial natural language terms to controlled vocabularies in various database systems. In the automatic and online mapping process, statistics and intelligent computer algorithms are employed to provide weighting for linking user's initial terms to controlled vocabularies ranked by their relevance. Users can then enter the controlled vocabularies to the information retrieval system for accessing the required items seamlessly and effectively. The success of this project depends very much on the quality of the mapping process, which seems to be an information retrieval system itself. Moreover, an additional layer, which can be as complicated as retrieving the actual item in the database system, is added in the subject access process. Nevertheless, this project demonstrates the feasibility of employing both natu-

ral language and controlled vocabulary in an information retrieval system.

Howarth (2004b) also discusses the development of a natural language gateway to metadata-enabled repositories. In order to combine the strengths of natural language and controlled vocabulary, "a common, natural language 'gateway' that could enable end-users to access resources without having to know the underlying metadata tagging, or domain-specific vocabularies" is required (Howarth, 2004b, 61). This gateway would directly link users to resources, eliminating the additional layers for end-user's convenience. Issues of multiple languages, quality assurance and interoperability in the development of "subject gateways" are discussed in Day and Neuroth (2004), Freyre and Naudi (2003), Howarth (2004a), Kelly, Closier and Hiom (2005), Vizine-Goetz and Beall (2004) and Tramullas and Garrido (2005). However, as Howarth's (2004b) study concludes, some metadata fields seem to be problematic in terms of matching natural language and the metadata vocabularies. Further investigation is necessary to develop more useful terminology and matching mechanism.

2.2.3 *From user terms to expertise terms*

In response to the new searching capabilities enabled by a computer system, Bates (1986, 1998) proposes the implementation of an end-user thesaurus that provides indexes from user's terms to controlled terms. This thesaurus would enable the information retrieval system to fulfill the design principles that Bates (1986) conceives in the online environment, namely: 1) indexing and access according to the "uncertainty principle;" 2) high variety or redundancy principle; and, 3) complexity principle. These principles are developed from the recognition that, while the document is a representation of a definite state of knowledge, a query related to information need appears to be a representation of an anomalous, inadequate and incoherent state of knowledge. Recently, a number of commercial information retrieval systems have implemented a thesaurus-enhanced interface (Shiri, Revie and Chowdhury 2002).

Green and Fraser (2004) conduct an empirical investigation of the semantic relationship between approximately 600 English verbs that have multiple meanings. Their study aims to raise the awareness of verb polysemy among people involved in using or developing thesauri. Spiteri (2002, 24) enriches the hierarchical displays in information retrieval thesauri by proposing a theoretical framework of incorporating

34

Knowl. Org. 37(2010)No.1
J. H. Leong. The Convergence of Metadata and Bibliographic Control?

word association testing in the development of the-sauri. This framework enables thesaurus developers to:

a) compile a list of terms that end-users associate most frequently with a chosen descriptor;
b) determine how end-users inter-relate response and stimulus terms; and,
c) incorporate user-defined term relationships within thesaurus displays.

Zeng and Chen (2003) discuss the development of an integrated thesaurus management and cross-thesaurus search system for facilitating interoperability among networked database systems. This integrated system includes controlled subject headings and thesauri of different schemas. It demonstrates how the usage of controlled vocabulary and thesauri synthesize individual databases, transcends the limitation of the type of resource in a database, and maintains dynamic updates of indexes to data.

Interestingly, when traditional catalogue systems are digitized and networked, they become online database systems that can also benefit from the aforementioned thesaurus projects. The following section details the challenges and opportunities generated by the transition of catalogue systems from the separated and manual environment to the networked and automated world.

2.3 Digitizing and networking the traditional catalogue systems

In terms of subject access, traditional catalogue systems focus on Cutter's two objectives of subject approach (Cutter 1904, 12): "1) to enable a person to find a book of which the subject is known and 2) to show what the library has on a given subject." As a result of unique and uniform subject headings, identification, retrieval, and browsing of library materials have been precise, complete and relevant in their ideal practice. In other words, catalogue systems that use subject heading and classification systems in subject access yield a rather high performance in the measures of precision and recall (Gorman 2004).

Library catalogue systems were designed originally as manual systems for handling traditional library materials that are tangible, well-defined, self-contained and relatively stable. Without modifications, catalogues cannot utilize the new capabilities of automated and networked systems. As many principles of catalogue systems originate from a manual card catalogue approach, having catalogue systems networked

and digitized entails a review of these principles. For example, the traditional cataloguing system's definition of access points, usually by title, author and subject, becomes less critical because an automated catalogue system enables every single field in the catalogue record to become an access point. In other words, subject access is not limited to the controlled headings in the subject fields. It can be enriched by including keywords in title, author, notes and other identifiable fields in the catalogue records. This enrichment seems to be the precursor of the cooperation model between natural language and controlled vocabulary.

2.3.1 From card to computer–the golden opportunity of online catalogues

The replacement of card catalogues by online catalogue systems in the 1970s-1980s raises some far-reaching and important problems. As Hildreth (1985, 272) pointed out, although online searches are popular and satisfying to users, "closer analysis of user search sessions … reveals that many such searches fail … or do not retrieve all relevant materials." Early online catalogues brought with them a few shortcomings, such as the lack of maintenance and loss of syndetic structure. Most importantly, catalogue cards, over years of being used and maintained, captured many interesting sorts of data, including dashed-on entries, local data and even dirt (an indicator of an item's popularity). When they were replaced by electronic files, these data were lost (Baker 1994). Nevertheless, these shortcomings have gradually been overcome as technology continues to advance.

The benefits of an online catalogue surpass its problems. For instance, an online catalogue not only indexes every field of the catalogue record, but also locates any words in a particular field regardless of the order of their appearance. In the manual system, only the beginning word of the pre-defined access points can be searched systematically. An automated catalogue system is able to index and provide access to every single field in a catalogue record. In addition, with the new capabilities of Boolean and proximity operators, users can have almost infinite searching strategies to formulate their queries and search for desired items. Subject access has been improved by using a combination of controlled vocabulary, natural language, and searching operators. In other words, an online catalogue system has the potential to deconstruct the dichotomy between approaches of hierarchical browsing, facilitated by controlled vocabulary and subject authority files, and analytical searching,

which is enabled by the capabilities for locating and combining character strings anywhere in the catalogue record (Dodd 1996). An online catalogue system with these capabilities can then combine the features and strengths of hierarchical browsing and analytical searching for providing more seamless subject access.

When the searching capabilities of an online library catalogue system are fully realized, users may feel overwhelmed by the numerous "combinatorial choices of subject elements" (Bates 1989, 405). Users also need assistance in adjusting the size of output sets according to their information needs. Bates suggests the implementation of an end-user super-thesaurus that incorporates vocabulary for online search features and capabilities (1989). The importance of controlled vocabulary and the authority control structure in the catalogue system is reinforced in the electronic environment (Gross and Taylor 2005). Beside their traditional role of maintaining the internal structure of the catalogue records, controlled vocabularies and their authority structure provide the linkages and hierarchies to develop this super-thesaurus. Moreover, there are presumably users who prefer a compact and precise method of information retrieval enabled by a controlled vocabulary system in academic research.

2.3.2 The paradise lost of online catalogues

Among many other good suggestions to enhance the capability of online catalogues, the concept of super-thesaurus has not been implemented in most Online Public Access Catalogues (OPACs), which often adopt the Machine Readable Cataloging (MARC) standard for providing interoperability in the networked environment. Apart from making the catalogue systems available online and including a few more access points, many OPACs remain relatively unchanged in terms of information retrieval theory and practice. They often assume a highly structured database and expert searchers who are familiar with their conceptual framework. In their study to compare the use of OPACs and Google™, Campbell and Fast (2004) argue that while Internet search engines assist novice users by giving them starting points so that they can develop better knowledge of these systems, OPACs generally fail to direct users' common entry terms to the access points that exist in their systems. Although both OPACs and Internet search engine systems are recognized as having similar goals, Campbell and Fast (2004) conclude that some changes are necessary for making OPACs engaging systems to users who are used to turning to the Web for fulfilling

their information needs. Implementing an end-user thesaurus that links users' entry terms to the controlled terms in the systems can facilitate these changes. On the one hand, OPACs can build on the strengths of traditional catalogue systems that provide a rich cross-reference network in the form of authority file systems and the hierarchies of subject heading references, such as related terms, used-for terms, see also references, broader terms and narrower terms. On the other hand, some conventional practices used in OPACs, such as a limited number of access points and subject headings and enumerative classification systems, would need to be modified to fully utilize the capabilities of the online environment and to meet the different expectations of users in terms of subject access. Nevertheless, with appropriate modifications, OPACs have great potential to combine the strengths of both bibliographic control and metadata. Markey (2007), for example, suggests that the paradise lost for OPACs can be regained if they embrace post-Boolean probabilistic searching, subject cataloguing and qualification metadata. In Figure 1, the main applications and rules dimensions show that OPACs and MARC seem to be appropriate entities for the merging of bibliographic control practices and metadata standards.

2.3.3 The transformation of subject headings
in the digital age

Not all the potentials of the electronic catalogue system are fully realized because subject analysis theory and practice is still rooted in the manual card cataloguing environment. Most OPACs employ subject headings and classification notations that are developed from an enumerative approach. *Library of Congress Subject Headings* (*LCSH*), the indexing language most widely used in OPACs, is largely developed from an enumerative syntax (Svenonius 2000). In general, a subject heading or classification notation derived from an enumerative approach has been pre-established as a whole concept, which may indeed consist of a number of sub-concepts, providing the exact meaning according to the item's context. It requires not only a full-string manipulation in searching but also pre-coordination in establishing headings at the indexing and searching stages. In this approach, headings are combined at the stage of indexing or cataloguing. As a result, enumerative subject headings and classification systems cannot fully utilize the flexibility of combining subject keywords by Boolean or other conjoining operators and notational synthesis.

36

Knowl. Org. 37(2010)No.1
J. H. Leong. The Convergence of Metadata and Bibliographic Control?

2.3.3 The rise of faceted subject analysis

Recent online information retrieval studies have recognized the limitations of a pre-coordinated subject approach and have suggested the use of a faceted approach to constructing subject headings and classification notations. In classification literature, the faceted approach to subject analysis was first introduced by Ranganathan in his *Prolegomena to Classification* in 1933. In each basic class, Ranganathan (1962) considers a notation a combination of pre-defined facets (aspects), namely Personality, Matter, Energy, Space and Time (PMEST), expressed in that citation order. Because of the lack of universal application of the PMEST facets and the complexity of building the faceted notations accordingly, Ranganathan's *Colon Classification* had not been very popular outside India. Nevertheless, the faceted classification approach demonstrated in *Colon Classification* embodies the computer amenability that is needed in managing and organizing information in the digital and networked environment.

Faceted subject analysis generates discrete units of classification categories. Using Ranganathan's theory, for example, a subject category can be comprised of finer sub-categories in the pre-defined PMEST and the basic class categories. Significantly, computer systems can operate and manipulate more efficiently and effectively on these discrete units. Subject terms in each facet can be joined or disjoined in the process of searching. Term coordination can happen at any stage in the cataloguing or retrieval process, which include "1) during vocabulary construction; 2) at the stage of cataloging or indexing; or, 3) at the point of retrieval" (Chan 2000). A faceted approach appears to be more flexible and simpler in syntax. Faceted headings can be combined in any way as needed. Complex headings can be broken down into different facets according to established categories. A faceted approach, with its post-coordinate practice, simpler syntax and discrete basic unit, may enable major subject heading and classification systems to benefit from the capability of computer systems so that they can better adapt to the digital environment. La Barre (2007), for example, argues that faceted classification might bridge the gap between library and Web communities. Her thorough review of faceted applications indicates that the Web community has been using the faceted practice without realizing that there are already faceted systems developed in library and information science literature. Meanwhile, the library community may not be familiar with the computing technology and concepts that

can facilitate the implementation of a faceted approach.

Bean and Green (2003), using the term "frame representations"—a facet in a faceted subject notation, suggest that the user requirement of high recall and high precision can be achieved by adopting frame representations in developing subject retrieval systems. Query structure, information visualization and the overall information retrieval process can be enhanced by applying the faceted approach to knowledge organization (Binding and Tudhope 2004). Recognizing the logical, semantic and syntactic strengths of Ranganathan's faceted analysis theory and the Classification Research Group's enrichment of the faceted approach, Broughton (2002) enhances and proposes the use of *Bliss Bibliographic Classification* (*BC2*), a faceted classification system, to manage Internet resources. Crowston and Kwasnik (2004), realizing the advantages of faceted classification, adopt this approach to address the issues of multidimensionality of classifying genres. In their discussion of using document genres as a means to retrieve useful information, Crowston and Kwasnik argue that a document's genre is a subtle and complex concept. The content and form of a document are often intertwined with its purpose and function. Consequently, the classification of genres appears to be a multidimensional phenomenon that can be addressed by using a faceted classification approach.

Spiteri (1998) highlights the ability to express compound subjects in faceted classification systems, and this ability has become a critical characteristic in online information retrieval systems. The faceted subject approach has produced significant impact in the traditional classification community. In their strategic plan to adapt to the new environment, the Subcommittee on Metadata and Subject Analysis of the Subject Analysis Committee of the Association of Library Collections and Technical Services (1999) has made a number of suggestions. One of them is to adopt a faceted version of *LCSH* in the subject fields in order to provide more satisfactory subject access to digital objects (Chan 2005). The faceted *LCSH*, known as Faceted Application of Subject Terminology (FAST), aims to "minimize the need to construct complex subject headings" by simplifying the syntax, adopting a faceted approach, and retaining the richness of the *LCSH* vocabulary (Chan 2005, 415). The FAST development demonstrates the efforts and initiatives of subject heading and classification systems to adapt to the new information development. It can facilitate the cooperation between the bibliographic control and

metadata approaches by creating a subject heading system that can be adopted by both approaches. The use of FAST headings in the subject metadata fields allows an information retrieval system to benefit both from the strengths of metadata and bibliographic control. Nevertheless, issues of consistency and continuity have to be addressed for existing information retrieval systems to adopt FAST headings. In the second part of the following section, the issues surrounding the implementation and applications of FAST by OCLC will be discussed in more detail.

2.4 Electronic publishing and the Internet

The tremendous volume of electronic publications available in the Internet creates unprecedented challenges to information organization and retrieval. Moreover, many of these digital materials appear to be amorphous, ill-defined, not self-contained, unstable and volatile. These characteristics of electronic resources have initiated dramatic changes in traditional subject access theory and practice. Consequently, metadata schemas have been developed by various domain-specific communities to handle the issues of organizing and providing access to online information resources (Campbell 2002; Greenberg and Campbell 2003; Tramullas and Garrido 2005). However, as illustrated in the **objects** dimension of Figure 1, collaboration between metadata schemas and bibliographic control is required to deal with the overlapping entities of electronic resources and the so-called "traditional" resources, such as electronic books and journals. This leads to the investigation of modifying a traditional subject analysis approach on the side of bibliographic control to enable the cooperation between metadata schemas and bibliographic control.

3.0 A comparison of metadata schemas and traditional catalogue systems

Metadata schemas and traditional catalogue systems share similar objectives in terms of providing subject access to information resources. Nevertheless, distinctive features can also be highlighted in these two systems. In terms of subject access, some metadata, such as folksonomies and machine-generated indexes, primarily use natural language or keywords for the sake of efficiency and simplicity (Guy and Tonkin 2006). This practice is complemented by statistical tools and computer algorithms in order to improve the measures of precision and recall of the system. Meanwhile, traditional catalogue systems often adopt a controlled

vocabulary, which requires trained experts in subject analysis. Metadata schemas, supported by automated metadata generation mechanisms, require less time and fewer human resources to generate metadata records (Campbell 2002; Greenberg, Sutton and Campbell 2003). Created primarily for digital resources, these records can fully utilize the capability of computer applications. This utilization of computer power enables metadata schemas to better accommodate a high quantity of objects. Appearing to be very standardized and consistent, catalogue systems using the bibliographic control approach usually provide quality records that yield a high performance in terms of precision and recall (Chan 2000).

3.1 A spiral dance between bibliographic control and metadata schemas –the automation of subject analysis

Two major strategies for adapting bibliographic control techniques to the application of metadata schemas are developed. The first strategy would be the automation of the assignment of subject headings and classification notations, which, once automated, should be integrated. In terms of automatic classification development, *DDC* seems to demonstrate a greater flexibility than LCC. For example, OCLC has developed a system that enables the browsing and searching of a *DDC* notation for a subject automatically. In addition, automation of *DDC* has also been successfully incorporated in the NetFirst, CyberDewey and Cooperative Online Resource Catalog (CORC) projects (Chan 2000; Vizine-Goetz 1997; Vizine-Goetz and Beall 2004).

The automation of constructing LCC notations seems to have started drawing researchers' attention. For example, the design of a hierarchical interface (HI) to LCC (HILCC) has been proposed by Davis (2002). This interface provides automatic subject access to resources in the libraries' electronic collections with a mapping table of three levels that links OPAC records against the HILCC. This system enables the browsing of subject menus and may serve as a model for the automation of the assignment of LCC according to established subject headings. Frank and Paynter (2004) also developed a feasible model of automating the assignment of LCC notations based on established subject headings. Their model primarily uses a tree-nodes approach to organizing LCC notations and *LCSH* terms. An LCC notation is represented as the node and *LCSH* as the branches. The heading at the deepest level of the tree would be the most specific

38

Knowl. Org. 37(2010)No.1
J. H. Leong. The Convergence of Metadata and Bibliographic Control?

TK ELECTRICAL ENGINEERING. ELECTRONICS. NUCLEAR ENGINEERING
TK7800-8360   Electronics
TK7880-8360   Applications of electronics
TK7885-7895   Computer engineering. Computer hardware
TK7888.3   Digital computers
TK7888.4   Circuits

Figure 2. Hierarchy of LCC schedules illustrating the concept of "circuits" (Williamson 1986, 47)

heading within the LCC given in the node. These initiatives demonstrate the possibility of automating the assignment of LCC numbers.

Some of the issues and the complexity of this process were discussed in Williamson (1986), who explored the feasibility of automating the LCC system for the online environment. She raised four primary problems that would need to be addressed in the computerization of the LCC system. The first issue of converting printed schedules to machine-readable form has been resolved. The following three problems continue to hinder the use of LCC in the digital environment. First, LCC's high degree of enumeration entails complex schedule layouts and topic displays. For example, the automated LCC system will need to display the whole hierarchy in Figure 2 in order to illustrate the complete meaning of the concept "circuits" in LCC's schedules. To further complicate the issue, there is also "hidden" enumeration from the "use of auxiliary tables and the 'A-Z' instructions for alphabetical arrangements using book numbers to create subtopics" (Williamson 1986, 47).

The lack of coordination, integration and mnemonics among LCC's classes, all of which were developed and have been perpetuated separately, created another problem for the automation of LCC. This situation creates obstacles for LCC's automation that requires consistency and coordination among its classes (Williamson 1986). The third problem in the automation of LCC is created by the non-hierarchical characteristic of LCC's notation. LCC's hierarchy is illustrated by its schedules that reflect classes, subclasses and constituent subjects, such as the one in Figure 1, rather than its notations. All these obstacles can be overcome eventually with more human effort and financial funding. Indeed, the recent implementation of Cataloguer's Desktop and Classification Web by the Library of Congress demonstrates that substantial efforts have been made in LCC's automation. Nonetheless, the issues raised by Williamson remain hindrances to the seamless adoption of LCC in an online environment.

3.2 A spiral dance between bibliographic control and metadata schemas—faceted subject headings

The second strategy attempts to simplify the subject headings and incorporate the faceted approach in order to enable the exploitation of bibliographic control in metadata applications and to address the complexity and computer amenability issues. A faceted heading is comprised of multiple concepts structured according to the pre-defined facets, which seem to be simpler in syntax. Moreover, the discreteness of the facets can make better use of the computer capabilities. The contrast between faceted and enumerative subject analysis has been illustrated in the section discussing the challenges created by digitizing and networking catalogue systems. This section therefore focuses on a particular project that transforms the *LCSH* from the enumerative approach to the faceted approach. Initiated by OCLC, the Faceted Application of Subject Terminology (OCLC FAST) project aims to exploit the strengths of the controlled vocabulary of *LCSH* and the simplicity and computer amenability of faceted subject headings for its adoption in metadata. OCLC FAST is "derived from *LCSH* but applied with simpler syntax" and adopts a faceted approach so as to retain the "richness of the *LCSH* vocabulary" for metadata applications (Chan 2005, 415). Developed from the existing headings that have already been used, the following facets have been conceived: topical, geographic, form, period, personal names, corporate names, conference/meetings, uniform titles, and name-title entries. Moreover, all OCLC FAST headings, except for the period facet, have been established in an OCLC FAST authority file. Although specificity may have been lost in providing more flexibility and simplicity, the goal of developing a new subject schema for metadata that builds on the rich vocabulary of *LCSH* while being easy to use, understand and maintain seems promising (Chan 2005).

OCLC FAST is a relatively new project, its success and adoption in the metadata environment are to be evaluated and determined. Several studies have been

published about the implementation and applications of OCLC FAST. Mitchell and Hsieh-Yee (2007) conduct a study to assess the feasibility of applying OCLC FAST headings to records in the *Ulrich's Periodicals Directory™* for providing subject access. They find that converting the sample of 100 random Ulrich's subject terms to OCLC FAST headings is useful and fairly easy to accomplish. Acknowledging a few problems that are resolved fairly easily, the authors conclude that the adoption of OCLC FAST headings would benefit the Ulrich's system as it will provide an addition means of subject access. Although the actual incorporation of OCLC FAST terms in Ulrich's records has yet to come, this feasibility study has demonstrated the possible adoption of a faceted *LCSH* by other subject classification systems.

The OCLC FAST project has received more attention, particularly in the digital environment. Several libraries have used OCLC FAST terms to provide subject headings for their digital projects. Nevertheless, several issues needed to be addressed in this project. First, the authority of these headings is established by the literary warrant of OCLC's WorldCat records, which happen to be records of primarily non-digital materials. These headings may not be able to describe adequately some of the content of digital and networked resources that often convey new knowledge. Second, the headings remain pre-coordinated within the same facet. The topical headings, for example, usually contain a number of interesting sub-topics, such as Revolution (United States, 1775-1783) and Quartets (Pianos (2), percussion). This multiplicity in the topical facet recreates issues of using the enumerative approach, undermining the project's efforts to achieve simplicity and computer amenability. Last but not least, the decline in quality and accuracy of OCLC FAST, a simplified version of *LCSH*, remains to be assessed. Will users and cataloguers accept the decline in quality and accuracy? In a qualitative analysis of 5000 bibliographic records converted from *LCSH* to OCLC FAST, the quality and accuracy issues are highlighted. Sponsored by ALCTS Subject Analysis Committee and reported by Qiang Jin (2008), the analysis concludes that although most OCLC FAST headings cover the "aboutness" in post-coordination, some headings become generic and lose their meanings. Relationship established in enumerative headings is lost in post-coordinated headings. Some OCLC FAST headings in certain facets make little or no sense when they are taken out of context.

In addition to the issues discussed above, OCLC FAST headings are often criticized for its partially fac-

eted syntax. OCLC FAST's topical facet remains enumerative. The problem of inconsistent syntax rules associated with *LCSH* passes on to OCLC FAST headings. Anderson and Hofman (2006), therefore, propose a fully faceted syntax for *LCSH*. Using the facets developed in *BC2*, Anderson and Hofman (2006, 8) demonstrate workable rules to restructure *LCSH* from enumerative syntax to faceted syntax. The fully faceted *LCSH*, in the words of the authors, "offers solutions to three problems that confront the current *LSCH* system: 1) inconsistent syntax rules; 2) inability to create headings that are coextensive with the topic of a work; and 3) lack of effective displays for long lists of subdivisions under a single subject heading in OPACs and similar electronic displays." The proposed syntax by Anderson and Hofmann shows that *LCSH* can become fully faceted. The success of this proposal depends on further testing and whether cataloguers and users find the new system more useful than the current approach. Interestingly, the model suggested by Anderson and Hofmann illustrates another step that can be made by subject cataloguing systems to move closer to the convergence of traditional bibliographic control and metadata.

4.0 Conclusion

This investigation of the current issues in bibliographic control and metadata schemas concludes that resolutions of these issues suggest a tendency towards convergence of these two approaches. Tools developed and adopted in these two approaches appear to effectively complement each other. To address the challenges of the rapid advancement of new knowledge in both digital and non-digital materials, information retrieval systems need to exploit the timeliness and computer amenability of metadata and the consistency, quality and conciseness of bibliographic control. Moreover, digitized and networked catalogue systems have enabled traditional bibliographic tools to incorporate both natural language and controlled vocabulary in providing subject access, while the metadata realm is recognizing the advantages of using the controlled vocabulary approach.

The huge volume of digital and networked information resources creates unprecedented challenges to both bibliographic control and metadata applications. In their attempt to provide quality subject access to these resources, metadata schemas and applications have recommended the use of traditional tools in bibliographic control, such as major subject headings and classifications systems. The bibliographic control

40

Knowl. Org. 37(2010)No.1
J. H. Leong. The Convergence of Metadata and Bibliographic Control?

community also responds to the proliferation of online resources by revising the *International Standard Bibliographic Description for Computer Files (ISBD(CF))* and re-titling it *International Standard Bibliographic Description for Electronic Resources (ISBD(ER))*. Chapter 9 of the *Anglo-American Cataloguing Rules* has been consequently modified and renamed from "Computer Files" to "Electronic Resources" (Howarth 2005). Another significant attempt of adapting cataloguing tools in the networked and digital environment is demonstrated by the FAST project, which revises *LCSH* using the faceted approach, aiming to provide an subject heading schema that appear to be "easy to use, understand, and maintain" for metadata (Chan 2005, 415).

Mitchell (1998) suggests that the differences between metadata schemas and bibliographic control may have been caused by different terminology used for similar concepts. Howarth (2005, 37) further argues the similarities of metadata and bibliographic approaches "have become sufficient to warrant a confluence in terminology and definition." This author has examined the trend and pattern of recent projects in bibliographic control and metadata for dealing with the issues and challenges created by the advancement of new knowledge, the choice of adopting natural language or controlled vocabulary approaches, the digitized and online catalogue systems and the proliferation of digital and online resources. The trends and patterns consequently identified suggest a larger convergence between the tools of bibliographic control and that of metadata. This convergence will lead to the triumph of the hybrid approach, a combination of the human approach of control vocabulary and the automation approach of algorithmic generation of metadata, in providing subject access. The dichotomy of human and automation approaches has turned into a spiral, which reaches a higher level of subject access every time when these two approaches meet and complement each other.

References

ALCTS/CCS/SAC/Subcommittee on Metadata and Subject Analysis. 1999. *Subject data in the metadata record: recommendations and rationale: a report from the ALCTS/CCS/SAC/Subcommittee on metadata and subject analysis*. Retrieved May 5, 2006, from http://www.ala.org/ala/alctscontent/catalogingsection/catcommittees/subjectanalysis/metadataandsubje/subjectdata.htm.

Anderson, James and Hofmann, Melissa. 2006. A fully faceted syntax for *Library of Congress Subject Headings*. *Cataloging & classification quarterly* 43(1): 7-37.

Baker, Nicholson. 1994. Discards. *The New Yorker* April 4, 1994: 64-86.

Bates, Marcia. 1986. Subject access in online catalogs: a design model. *Journal of the American Society for Information Science* 37: 357-76.

Bates, Marcia. 1989. Rethinking subject cataloging in the online environment. *Library resources & technical services* 33: 400-12.

Bates, Marcia. 1998. Indexing and access for digital libraries and the internet: human, database, and domain factors. *Journal of the American Society for Information Science* 49: 1185-1205.

Bean, Carol and Green, Rebecca. 2003. Improving subject retrieval with frame representations. In McIlwaine, Ia ed., *Subject retrieval in a networked environment*: proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001. Munchen: K.G. Saur, pp. 114-21.

Beghtol, Clare. 2004. Naïve classification systems and the global information society. In McIlwaine, Ia ed., *Knowledge organization and the global information society: proceedings of the eighth international ISKO Conference, 13-16 July 2004, London, UK*. Würzburg: Ergon, pp. 19-22.

Binding, Ceri and Tudhope, Douglas. 2004. Integrating faceted structure into the search process. In McIlwaine, Ia ed., *Knowledge organization and the global information society: proceedings of the eighth international ISKO Conference, 13-16 July 2004, London, UK*. Würzburg: Ergon, pp. 67-72.

Broughton, Vanda. 2002. Faceted classification as a basis for knowledge organization in a digital environment: the *Bliss Bibliographic Classification* as a model for vocabulary management and the creation of multidimensional knowledge structures. *The new review of hypermedia and multimedia* 7: 67-102.

Buckland, Michael. 1999. Mapping entry vocabulary to unfamiliar metadata vocabularies. *D-Lib Magazine*, 5(1). Retrieved May 5, 2006, from http://www.dlib.org/dlib/january99/buckland/01buckland.html.

Campbell, D. Grant. 2002. The use of the Dublin Core in web annotation programs. *Proceedings: International Conference on Dublin Core and Metadata for e-Communities October 13-17, 2002 Florence, Italy*. Florence: Firenze University Press, pp. 105-10.

Knowl. Org. 37(2010)No.1
J. H. Leong. The Convergence of Metadata and Bibliographic Control?

41

Campbell, D. Grant and Fast, Karl V. 2004. Panizzi, Lubetzky, and Google: how the modern web environment is reinventing the theory of cataloguing. *Canadian journal of information and library science* 28: 25-38.

Cataloging Policy & Support Office. 2006. Library of Congress subject headings weekly lists. Retrieved July 4, 2006, from http://www.loc.gov/catdir/cpso/.

Chan, Lois. 1994. *Cataloging and classification: an introduction*. New York: McGraw-Hill.

Chan, Lois. 2000. Exploiting *LCSH*, LCC, and *DDC* to retrieve networked resources: issues and challenges. Retrieved Mar. 8, 2006, from http://www.loc.gov/catdir/bibcontrol/chan_paper.html.

Chan, Lois. 2005. *Library of Congress Subject Headings: principles and application*. Westport, Conn.: Libraries Unlimited.

Crowston, Kevin and Kwasnik, Barbara. 2004. *A framework for creating a facetted classification for genres: addressing issues of multidimensionality*. Retrieved June 7, 2006, from http://crowston.syr.edu/papers/hicss04genre.pdf.

Cutter, Charles. 1904. *Rules for a dictionary catalog*. Washington: Government Printing Office.

Davis, Stephen. 2002. HILCC: A hierarchical interface to Library of Congress Classification. *Journal of internet cataloging* 5(4): 19-49.

Dodd, David. 1996. Grass-roots cataloging and classification: food for thought from world wide web subject-oriented hierarchical lists. *Library resources & technical services* 40: 275-86.

Frank, Eibe and Paynter, Gordon. 2004. Predicting library of congress classifications from library of congress subject headings. *Journal of the American Society for Information Science and Technology* 55: 214 – 27.

Freyre, Elisabeth and Naudi, Max. 2003. MACS: Subject access across languages and networks. In McIlwaine, Ia ed., *Subject retrieval in a networked environment*: *proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001*. Munchen: K.G. Saur, pp. 3-10.

Gill, Tony. 2000. Metadata and the world wide web. In Baca, Murtha ed., *Introduction to metadata: pathways to digital information*.

Gorman, Michael. 2004. Authority control in the context of bibliographic control in the electronic environment. *Cataloging & classification quarterly* 38 (3/4): 11-22.

Green, Rebecca and Fraser, Lydia. 2004. Patterns in verbal polysemy. In McIlwaine, Ia ed., *Knowledge organization and the global information society: proceedings of the eighth international ISKO Conference, 13-16 July 2004, London, UK*. Würzburg: Ergon, pp. 29-34.

Greenberg, Jane, Sutton, Stuart and Campbell Grant. 2003. Metadata: A fundamental component of the semantic web. *Bulletin of the American Society for Information Science and Technology* 29(4): 16-18.

Gross, Tina and Taylor, Arlene. 2005. What have we got to lose? The effect of controlled vocabulary on keyword searching results. *College & research libraries* 66: 212-30.

Guy, Marieke and Tonkin, Emma. 2006. Folksonomies: Tidying up tags? *D-Lib Magazine,* 12(1). Retrieved 28 Feb. 2006, from http://www.dlib.org/dlib/january06/guy/01guy.html.

Hildreth, Charles. 1985. Online public access catalogs. *Annual review of information science and technology* 20: 233-85.

Howarth, Lynne. 2004a. Metadata schemas for subject gateways. *International cataloguing and bibliographic control* 33: 8-12.

Howarth, Lynne. 2004b. Modelling a natural language gateway to metadata-enabled resources. In McIlwaine, Ia ed., *Knowledge organization and the global information society: proceedings of the eighth International ISKO Conference, 13-16 July 2004, London, UK*. Würzburg: Ergon, pp. 61-66.

Howarth, Lynne. 2005. Metadata and bibliographic control: soul-mates or two solitudes?. *Cataloging & classification quarterly* 40(3/4): 37-56.

Jin, Qiang. 2008. Is FAST the right direction for a new system of subject cataloging and metadata? *Cataloging & classification quarterly* 45(3): 91-110.

Kelly, Brian, Closier, Amanda and Hiom, Debra. 2005. Gateway standardization: A quality assurance framework for metadata. *Library trends,* 53(4): 637–66.

Koch, Traugott. 2000. Quality-controlled subject gateways: definitions, typologies, empirical overview. *Online information review* 24(1): 24 – 34.

La Barre, Kathryn. 2007. Faceted navigation and browsing features in new OPACs: a more robust solution to problems of information seekers? *Knowledge organization* 34: 78-90.

Lee-Smeltzer, Kuang-Hwei. 2000. Finding the needle: Controlled vocabularies, resource discovery, and Dublin Core. *Library collections, acquisitions & technical services* 24: 205-15.

Markey, Karen. 2007. The online library catalog: paradise lost and paradise regained? *D-Lib Magazine,* 13(1/2). Retrieved October 24, 2008, from

42

Knowl. Org. 37(2010)No.1
J. H. Leong. The Convergence of Metadata and Bibliographic Control?

http://www.dlib.org/dlib/january07/markey/01markey.html.

McIlwaine, Ia. 2000. Interdisciplinary: a new retrieval problem? In Beghtol, Clare, Howarth, Lynne and Williamson, Nancy ed., *Dynamism and stability in knowledge organization: proceedings of the 6th International ISKO Conference, Toronto, Canada, 10-13 July 2000*. Wurzburg: ERGON, pp. 261-67.

Mitchell, Joan. 1998. In this age of WWW is classification redundant? *Catalogue & index* 127: 5.

Mitchell, Vanessa and Hsieh-Yee, Ingrid. 2007. Converting Ulrich's™ subjecting headings to FAST headings: A feasibility study. *Cataloging & classification quarterly* 45(1): 59-85.

Ranganathan, Shiyali. 1962. *Elements of library classification*. Bombay: Asia Publishing House.

Rowley, Jennifer. 1994. The controlled versus natural indexing languages debate revisited: a perspective on information retrieval practice and research. *Journal of information science* 20: 108-19.

Shiri, Ali, Revie, Crawford and Chowdhury, Gobinda. 2002. Thesaurus-enhanced search interfaces. *Journal of information science* 28: 111-22.

Spiteri, Louise. 1998. A simplified model for facet analysis: Ranganathan 101. *Canadian journal of information and library science* 23: 1-30.

Spiteri, Louise. 2002. Word association testing and thesaurus construction: defining inter-term relationships. In Howarth, Lynne, Cronin, Christopher and Slawek, Anna ed., *Advancing knowledge: expanding horizons for information science: proceedings of the 30th Annual Conference of the Canadian Association for Information Science 30 May – 01 June, 2002.* Toronto: CAIS, pp. 24-33.

Svenonius, Elaine. 2000. *LCSH*: semantics, syntax and specificity. *Cataloging & classification quarterly* 29(1/2): 17-30.

Tillotson, Joy. 1995. Is keyword searching the answer? *College & research libraries* 56: 199-206.

Tramullas, Jesús and Garrido, Piedad. 2005. Constructing web subject gateways using Dublin Core, RDF and topic maps. *Information research*, 11(2). Retrieved May 6, 2006, from http://InformationR.net/ir/11-2/paper248.html.

Vizine-Goetz, Diane. 1997. From book classification to knowledge organization: Improving internet resource description and discovery. *Bulletin of the American Society for Information Science* 24(1): 24-27.

Vizine-Goetz, Diane and Beall, Julianne. 2004. Using literary warrant to define a version of the DDC for automated classification services. In McIlwaine, Ia ed., *Knowledge organization and the global information society. proceedings of the eighth international ISKO Conference.* Würzburg (Germany): Ergon Verlag, pp. 147-52.

Williamson, Nancy. 1986. The Library of Congress classification: problems and prospects in online retrieval. *International cataloguing* 15: 45-48

Zeng, Marcia and Chen, Yu. 2003. Features of an integrated thesaurus management and search system for the networked environment. In McIlwaine, Ia ed., *Subject retrieval in a networked environment*: *proceedings of the IFLA Satellite Meeting held in Dublin, OH, 14-16 August 2001*. Munchen: K.G. Saur, pp. 122-28.