PNLA QUARTERLY

The Official Journal of the Pacific Northwest Library Association

Volume 74, number 3 (Spring 2010)

Volume 74, number 3 (Spring 2010)

Feature Articles

Call for submissions and instructions for authors

Authors should include a 100-word biography and mailing address with their submissions. Submit feature articles of approximately 1,000-6,000 words on any topic in librarianship or a related field. Issue deadlines are October 1 (Fall), January 1 (Winter), April 1 (Spring), and July 1 (Summer). Please email submissions to mbolin2@unl.edu in rtf or doc format.

President's Message

SAMANTHA HINES

Spring brings warmer weather, budding trees, blooming flowers, and library conferences galore. I have been enjoying visits to many of the state and provincial conferences lately—It is great to get a chance to see the members of PNLA outside the PNLA conference, and realize how well connected we are as a region. I have always seen old friends at each conference and made new ones, and it has been a wonderful experience.

For those who have not been able to catch my spiel at your membership meetings, here are the highlights:

- We hope to see you at our 101 st Conference, to be held in Victoria BC at the Empress Hotel, August 11-14. Rooms are going fast so book now! Registration will be up soon (perhaps before you even see this message) so check pnla.org often. The preliminary program is available and it looks great! This is a joint conference with Washington Library Association, and should be very educational and also fun.
- The conference in 2011 will be held in downtown Spokane, WA, very convenient to shopping and attractions. 2012 will be held in Anchorage, Alaska. I scouted out hotel sites and local attractions at the recent Alaska Library Association conference and attendees in 2012 will be in for a treat!
- We are looking for applicants for LEADS this fall, to be held in Schweitzer, Idaho on October 17-22. Those who attend say the experience is one of the most valuable of their careers. Check out our website for more information.
- The ballot for the Young Readers' Choice Award in 2011 has just been announced! See our website for details.
- This fall we are hoping to launch some online education opportunities. We've always had high-quality programming at our conferences in addition to the great social and networking atmosphere. We will now be working to offer the best of our conferences online. Look for more details soon.
- Our new online membership software is up and running! You should be receiving notices and information now via your email. Please notify me or our Second Vice President, Jason Openo, if you are not.

It is an exciting time to be involved in PNLA as we start our second century! And, as always, I would love to hear from the members of PNLA as we proceed onward with new initiatives and new experiences. Please feel welcome to email me at Samantha.hines@umontana.edu or call me at 406-243-4558, or catch me at your local conference or at our annual meeting in Victoria. Thanks!

From the Editor

MARY BOLIN

The PNLA Board has made the decision that the *Quarterly* will be a peer-reviewed journal. This is a great time to make such a decision. Librarians in North America and elsewhere in the world are looking for peer-reviewed publication outlets. The *PNLA Quarterly* is receiving more submissions from librarians in Africa, India, and other countries. It has always been a high-quality journal, and this will enhance its quality further. It will take some time to get peer review set up and get a pipeline of articles into the process and ready for publication. As many journals do, we will continue to publish some things that are not peer-reviewed, but which are well-written and of interest to our readers.

Who wants to serve as a peer reviewer? Please let me know if you are interested. You can email me at mbolin2@unl.edu.

Characteristics and Choice of Librarianship as a Career among Librarians in Ogun State, Nigeria

M.O. Salaam

K.A. Owolabi

Both authors are librarians at Nimbe Adedipe Library University of Agriculture, Abeokuta, Ogun State, Nigeria. M.O. Salaam can be reached at: reachabdel@yahoo.com. K.A. Owolabi can be reached at: yomiowolabi2000@yahoo.com

Introduction

The library is the nerve center of academic institutions in Nigeria, but libraries and librarianship have evolved gradually in Nigeria and have not yet taken their proper place among the institutions and professions in the country. In spite of the vast and diverse knowledge and information resources available in libraries, they still remain the least popular institution and the least sought profession in Nigeria (Okoro, 2009) Given the fact that a strong academic background breeds career success, it is not difficult to understand why professional such as lawyers, doctors, engineers, or accountants are traditionally held in high esteem because they all demand solid academic backgrounds. There are some professions that require advanced training but never receive due respect, and librarianship is one of them. Librarianship is a profession that is dedicated to serving the public, to providing timely and accurate information, and which has a great impact to the development of the society. Nevertheless, the profession does not share the respect of those other professions mentioned above. Igbinosa (2007) maintains that students do not pursue careers in librarianship because it does not have the prestige given to other professions. This is in spite of the fact that librarians are the silent heroes behind the success stories of students, researchers, medical doctors, etc.

Literature Review

Nigerian youths are faced with problems in deciding on a career. Geshinde (1986), cited by Adeyemo (2003), maintains that there is a danger inherent in making wrong career choices, including adverse effect on health, difficulty making friends and finding colleagues, and lack of job security. Issa and Nwalo (2008) support this position by saying that the wrong choice of career can lead to frustration and low productivity.

Wiljers and Mijers (1996) define a career as an individual's life time of learning and work. Ferry (2006) highlights the factors that determine adolescent occupational choice, including life context, personal attitude, and educational attainment. Tella (2007) identifies factors that can promote the choice of a career in librarianship, including attractive salary, good conditions of service, and job security.

Agumanu (1989) conducted a studied on factors that influenced the students in Imo State University Library School to pick a career in librarianship. Eighty percent of

respondents agreed that they picked the profession when they failed to secure admission to other professions such as law and engineering, while 20 percent students chose the profession due to the influence of other librarians in their families. Alemna (1991) conducted a study discovered that most graduates surveyed chose librarianship after failing in other options. He also discovered that opportunity for intellectual development and ability to further education were reasons for choosing the profession. This corroborates the finding of Nzotta (1982), who discovered that a majority of respondents (58 percent) chose librarianship because it gives room for intellectual development.

Issa and Nwalo (2008) discovered that nearly 70 percent of respondents claimed chose the profession because of previous experience working in a library, while more than 15 percent saw librarianship as a means of job security.

Objectives of the Study

- To identify the reasons for the choice of librarianship as a profession
- To identify the sources of information used in choosing the profession
- To identify the level of satisfaction

Methodology

The study used a descriptive survey design and a questionnaire called Characteristics and Choice of Librarianship as a Career (CCLC). Descriptive statistical techniques such as frequency counts and percentages were employed in the analysis. The target population of the study was library personnel in eight libraries in Ogun State, which included five academic libraries, one public library, one special library, and the branch of the National Library in the state. The Purposive Sampling Technique (PST) was used to select 90 library personnel. The breakdown is as follows:

	Library and Institution	Population
1	Nimbe Adedipe Library, University of Agriculture Abeokuta, Ogun State	20
2	Federal College of Education, Abeokuta, Ogun State	15
3	Olabisi Onabanjo University, Ago- Iwoye, Ogun State	20
4	Moshood Abiola Polytechnic Abeokuta	06
5	Redeemers University, Mowe, Ogun State	08
6	Simeon Adebo Library Abeokuta	10
7	National Library of Nigeria Abeokuta	*10*
8	Neuropsychiatry Hospital, Abeokuta	*06*
	Total	*90*

Results and Discussion

Table 1: Educational attainment of the respondents

Academic Qualification	Frequency	Percentage
OND/DLS	23	25.6
BLS/BSC	12	13.3
MLS	53	58.9
PhD	02	1.8
Total	90	100

More than half of respondents have a Master's degree in Library Science (MLS).

Table 2: Reason for the choice of librarianship

Reasons	Frequency	Percentage
It gives room for intellectual development	23	26
Employment opportunity	19	21.1
Previous working experience in the libraries	33	37
Love for books and reading	06	07
Prospect for good salary	03	03
Attractive working environment	02	02
By accident	02	02

More than one-third of respondents chose librarianship because of their previous working experience.

Table 3: Time of choice

Time of choice	Frequency	Percentage
While working after secondary school.	43	48
While in the university	22	24
While working after NYSC	12	13
During NYSC	10	11
While in secondary school	03	03
Before attending secondary school	01	01
Total	90	100

Almost half of respondents developed an interest in the profession while working in the library after secondary school. This contradicts the finding of Nzotta (1982) that many people consider the profession during their first degree courses in universities.

Table 4: Sources of information used in choosing the profession

Source of information	Frequency	Percentage
Information from friends and relatives	41	45.5
Advertisements for vacancies for librarians	24	26.7
Library school prospective material	15	16.7
Information about librarianship as a career in radio, television, and newspapers	10	11.1
Total	90	100

Friends and relatives were the major sources of information in choosing the profession. This is in agreement with the finding of Agumanu (1989) that family members were the sources of information for students of Imo State University, Nigeria on the choice librarianship.

Table 5 Level of satisfaction

Level Of Satisfaction	Frequency	Percentage
Highly Satisfied	61	68
Satisfied	23	26
Not Satisfied	06	06

Nearly 70 percent of respondents were highly satisfied with librarianship as a profession.

Conclusion and Recommendations

Librarianship has not taken its rightful place among other professions. That fact that many librarians choose the profession because of their working experience in the library shows that there is not enough awareness among secondary school students who would

come afresh to the profession. This supports the findings of Nwalo and Isaac (2008) that working experience in the libraries prompts many Librarians in choice of librarianship as a profession. Many librarians who choose the profession after working in libraries after secondary schools do so to secure reliable jobs and promotions. There is not adequate information about the profession for those who would come to it through other channels. Many respondents chose their profession because there are other librarians in their families. A large number of respondents find satisfaction in their career.

The following recommendations are made based on the findings of the study.

- Library use should be introduced in all secondary schools in the country.
- Librarianship should be included in during career counseling for secondary school students.
- There is a need for special salary and allowances for librarians
- The Registration Council of Nigerian Library Association should engage in a massive campaign for awareness about librarianship as a noble profession.

References

Adeyemo, D. A. (2003). Career counseling for national development in education this millennium: Innovations in theory and practice. Bamisaye, O.A., Nwazubuike, I.A., & Okediran, A. (eds.). Ibadan.

Agunmanu, J. (1989). The choice of librarianship as a career in Nigeria. *Third World Libraries 1* (2)

Alemna, A.A. (1991). The characteristics of past postgraduates diploma students of the departments of library and archival studies, University of Ghana, Legion, 1981/ 82-1987/1988. *African Journal of Library, Archives, and Information Science 1* (1): 45-50.

Buddy, J.M. (1998). *The academic library: Its purpose and its operation.* Englewood, Colorado: Libraries Unlimited.

Ferry, N.M. (2006). Factors influencing career choice of adolescents and young adults in rural Pennsylvania. *Journal of Extension 44* (3) 3R IB7:1-6. Available: http://www.joe.org/2006june/rb75ht

Geshinde, S.A. (1986). Equipping students with career decision making skills. *Nigerian Journal of Counseling and Development 1* (1): 23-26.

Igbinosa, I.O. (2007). University of Benin students' perception of library and information science profession. *Nigerian Library Link 5* (1):74-82.

Issa, A.O., & Nwalo, K.I.N. (2008). Factors affecting the career choice of undergraduates in Nigerian library and information science schools. *African Journal of Library, Archives, and Information Science 18* (1): 23-31.

Montoya, L. M. (1999). Vital technical services in academic libraries. *Library Philosophy and Practice 2* (1). Spring: 1-15 Available: http://unllib.unl.edu/LPP/montoya.htm

Nzotta, B.C. (1982). Choice of librarianship as a career: The case of the post graduate class of 1978. *Nigerian Journal of Library and Information Studies 1* (1):1-11.

Okoro, C.C. (2009). Career choice of students of library and information science in tertiary institutions in Imo State, Nigeria. *The Information Technologist 6* (1) 53-59.

Tella, A., Ayeni S.O., & Popoola S.O. (2007). Work motivation, job satisfaction, and organization commitment of library personnel in academic and research libraries in Oyo State, Nigeria. *Library Philosophy and Practice.* Available: http://unllib.unl.edu/LPP/tella2.htm

Wijer, G.A., & Mijer, F. (1996). Career guidance in the knowledge society. *British Journal of Guidance and Counseling 24:* 185-198.

The Catalogers' Revenge: Unleashing the Semantic Web

Virginia Schilling

Virginia (Ginger) Schilling works at the Center for Bibliographic Studies & Research at the University of California at Riverside. She can be reached at: virginia.schilling@ucr.edu

Introduction

One theme that seems to be a consistent undertone in the library-related literature, and one that is reinforced by popular culture, is that cataloging, that meticulous process of identifying bits of data, is no longer necessary. Depending on the perspective, the tone varies between dread and glee. Cataloging is dead, they say. Catalogers are an anachronism and it is just a matter of time before that job function disappears entirely. We don't need catalogers; we have simple keyword searches and social tagging. We can just throw all that undifferentiated text out to the world and let the users sort it out. Look, these proponents say, how well it works on the World Wide Web. So I find it to be the supreme irony that the next revolution in web technology is slated to be exactly the opposite of this supposed trend away from the careful description and precise identification of "things" and the relationships between them.

The World Wide Web first revolutionized the presentation of text. People thousands of miles away from each other could suddenly see the same exact text or data at the same time in the same format. (Well, sort of. It was identical when there was only one browser. These days, of course, presentation of the text can vary between web browsers with different capabilities. Ask anyone who has tried to code a web page for both Microsoft's Internet Explorer and Mozilla's Firefox.) It is hard, now, to remember when it was pretty amazing just to be able to see web pages. The second wave of data handling has come with the collaboration technologies: social tagging, networking websites like Facebook and the interactivity of Wikis and blogs. Both of these technology sea changes have been aimed at

making data accessible to people. Both have improved a person's ability to read the text or data presented and interpret meaning from it, for good or for ill.

The next wave of technology will make data accessible to computers as well as people. Instead of undifferentiated text presented on a web page, each data point will be coded in a way that computer programs will be able to understand and interpret. This next wave of technology change will lead us into the semantic web. The World Wide Web Consortium (W3C) defines the semantic web as providing "a common framework that allows data to be shared and reused across application, enterprise, and community boundaries" (2009b, para 1). Campbell and Fast note "information will be machine understandable, as well as machine readable, enabling intelligent agents to draw sophisticated inferences from the metadata attached to Web-based information" (2004, p. 383).

Basic Structure of the Semantic Web

Web pages are generally coded using either HTML or the stricter XHTML markup languages (collectively known as X/HTML). However, these languages only tag data on the web page for presentation purposes (i.e. they say things like "make this word bold"), not for the actual meaning delivered by the content (they don't say "this word is the name of a city"). Using markup languages that code for meaning in addition to presentation will allow software to find and use specific bits of information on the web page, such as a date or a person's name, rather than just understanding everything on the page as one gigantic mass of text. Each bit becomes a separate piece of information with its own individual meaning. In some ways, the concept is like taking everything on the Internet and putting it into a gigantic distributed database.

The real power of semantic markup, however, is that implicit relationships between bits of data can be established by the computer. People can read the text of two different web pages, for example, and be able to interpret implicit relationships between the data in each one. A computer cannot do this. If on one web page, a city is stated to be in a particular country and on another separate web page, a person is stated to be in that same city, then the implicit statement that the person is located in that same country can be understood easily by a person. Semantic markup will allow that implicit relationship to be also understood by a computer.

Semantic web markup will consist of a vocabulary used within a defined syntax similar to the way that HTML is implemented. A software package designed to "understand" what the markup means will be able to extract and use the tagged information. For example, a search engine designed to "read" the tags that indicate a person's name versus those that indicate a corporate entity will be able to distinguish between a web page containing biographical information about the person Abraham Lincoln and a web page for an elementary school named Abraham Lincoln.

As shown in the example above, at the most general level, semantic tagging will improve the precision of search results for Internet search engines designed to use it. But beyond basic search results, it will also allow for things such as users tweaking the search engine, or agent as it is often called, to understand the user's context for searching. Harper and Tillett (2007) state:

Another large part of the Semantic Web vision is about enabling "agents" or systems to insert a searcher's/user's individual context or perspective into a search for information. This necessarily involves interacting with the elements that make up that context, such as

schedules, contacts, group membership, profession, role, interests, hobbies, location, etc. Systems can then be developed that "understand" the searcher's needs, based on who the searcher is and the searcher's "context" or demographics. (p. 65)

An overview of Current Research: Standards Development

Research in this field is led by the W3C. The actual technology to carry out the goals of the semantic web is still in its infancy, where it exists at all. Current research is being directed primarily towards establishing standards and developing basic specifications to ensure interoperability in the future and to allow the construction of the tools and components that will form the invisible backbone of the semantic web.

W3C has developed standards/specifications for an abstract model to describe relationships between "things", expressed as Resource Description Framework (RDF) (2004a), a semantic schema to allow the description of other vocabularies in RDF (RDFS) (2004b) and a syntax for RDF in XML (RDF/XML) (2004c). Gleaning Resource Descriptions from Dialects of Languages (GRDDL) is a specification for extracting RDF content from marked up XML or XHTML pages (2007). Simple Knowledge Organization System (SKOS) is a specification for converting existing controlled vocabularies into an RDF-compliant form (2009b). SPARQL Query Language for RDF is designed to do exactly what it says: query RDF-compliant data (2008c). The Web Ontology Language (OWL) is yet another extension of semantics for RDF, allowing for much more sophisticated use than that supported by the basic model and RDFS (2009c). RDFa is a specification for representing RDF in XML and XHTML documents (2008a). The specification for Protocol for Web Description Resources (POWDER) builds on these other specifications to allow the description of groups of web resources for purposes such as customized retrieval of resources or the identification of resource authenticity (2009a).

However, W3C is not the only organization developing tools and standards that will underpin the semantic web. There are organizations with projects contributing to development all over the world. Library of Congress has made its subject headings data available in RDF/XML (n.d.). The DCMI/RDA Task Group has started a project to convert Resource Description and Access (RDA) into RDF (2008). The International Federation of Library Associations and Institutions (IFLA) is busy translating its Functional Requirements for Bibliographic Records (FRBR) into RDF (2008).

Completely separate from W3C but with the same idea in mind, the open source community has developed a set of formats called "Microformats" (About microformats, n.d.). Just like RDFa, these allow the use of existing XHTML tags to add meaning to the data they mark up. hCalendar allows events to be tagged in such a way that the information can be extracted and, for example, added to a calendar somewhere else. hCard allows contact information to be marked up in the same way. Formats exist to describe resumes, reviews and Atom feeds. Other formats are under construction to describe audio, recipes and citations (Microformat, 2009). Talis provides another way of adding RDF-compliant tags to a web page with eRDF: Embeddable (or Embedded) RDF (Talis, 2006).

The National Archives in the United Kingdom has developed PRONOM, an authoritative registry of digital file formats for use in the RDF/XML environment (n.d.). The Global Digital Format Registry (GDFR), developed by Harvard (n.d.), is merging with PRONOM to become the UDFR or Unified Digital Formats Registry (2009). The Dublin Core Metadata Initiative (DCMI) has a registry of metadata schemes, The Dublin Core Metadata Registry (2008), as does the National Science Digital Library (NSDL). The NSDL Metadata

Registry "provides services to developers and consumers of controlled vocabularies and is one of the first production deployments of the RDF-based Semantic Web Community's Simple Knowledge Organization System (SKOS)" (2009, Welcome to The Registry! section). The JISC IE Metadata Schema Registry (IEMSR) "will act as the primary source for authoritative information about metadata schemas recommended by the JISC IE Standards framework" (2009, About IEMSR section).

There are numerous descriptive vocabularies that can be used for semantic markup of resources. DCMI designed Dublin Core (2005) specifically with web resources in mind. The Gateway to Educational Materials (GEM) describes web-based educational resources (2009). The Public Health Information Network (PHIN) vocabulary developed and maintained by the Centers for Disease Control and Prevention (CDC) "enables data from different programs to be consistently documented" (2005, p. 4). Hundreds of other vocabularies exist both within the focus of library work and completely outside of and unrelated to it: TEI, EAD, FOAF, DOAP and so on.

Finally, there is a myriad of various projects documented in the literature testing the possibilities of semantic web technology. Tonkin & Strelnikov (2009) discuss the JISC metadata registry mentioned above and Heery & Wagner (2002), the DCMI metadata registry. Hildebrand et al. (2009), Angjeli et al. (2009) and Guzmán Luna, Torres Pardo & López García (2006) each discuss projects to develop or use specific thesauri. Talantikite, Aissani & Boudjlida (2008), Arch-int & Sophatsathit (2003) and Uddin & Janecek (2007) all discuss the general use of ontologies. Chavarriaga & Macias (2009) look at modeling a semantic web-based interface. Damiani & Fugazza (2007) discuss the management of intellectual property rights using semantic web technologies.

Implementation and Use of the Technology

RDF is not used in web pages directly. It is simply a vocabulary that describes the relationship between two things. Just as data can be coded into XML and displayed by a program that reads XML, it can be coded into RDF and read by RDF-compatible software. RDF, however, cannot be read by standard web browsers at this time. Instead the web pages are marked up by some other method and used by a separate software program that can extract that information and translate it into RDF. Similar to the method used by OAIster, the metadata in the web page is harvested and stored by a separate tool. Currently, there are two basic methods for marking up data semantically in a web page.

The first, most basic, method is to use the existing meta tags available in the X/HTML header to directly code values for one or more descriptive metadata vocabularies, like Dublin Core. The second method is to extend existing X/HTML coding in the web page body using semantic tag attributes defined by one or more profiles, such as eRDF, RDFa or Microformats, to hold the metadata vocabularies. The choice of which method to use depends on the resource being described and for what purpose. Does a metadata harvester, for example, expect to find the information in meta tags or does it look for a link to a profile that defines the elements used in the body of the page?

This first method of using existing meta tags seems somewhat like a stop-gap measure to be used until better standards can supplant it. Meta tags don't identify data within the text in the body of the page. The data has to be pulled out and separately tagged in the head, meaning that if changed, it must be changed in two places, one of which may not be visible in WYSIWYG X/HTML editors. The functionality is also more limited than embedding the tags in the body because the metadata is often understood by RDF

extraction software to refer to the web page as a whole and not to individual pieces within the web page. In his brief discussion about embedding metadata in X/HTML, O'Donnell gives an example of embedded metadata. Adapted from his example, the use of Dublin Core in the meta tags might look something like:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">

<head>

<link rel="schema.dc" href="http://purl.org/dc/elements/1.1/" />

<meta name="DC.title" content=" Naked Metadata" />

<meta name="DC.creator" content=" Jonathan O'Donnell " />

<meta name="DC.rights" content=" http://purl.nla.gov.au/net/jod/tutorial/naked-
metadata.html © Jonathan O'Donnell 23 October 2005" />

<meta name="DC.date" content="23 October 2005" />

</head>

<body>

<h1>Naked Metadata</h1>

<h2>Jonathan O'Donnell</h2>

<p>http://purl.nla.gov.au/net/jod/tutorial/naked-metadata.html © Jonathan O'Donnell 23
October 2005</p>

</body>

</html>
```

The second method is the technique of using the existing X/HTML structure to include semantic tags within the body of a web page. In the head, a reference is made to a pre-existing profile that defines how the elements are used. Bits of data are enclosed with tags such as div, span or class. Formatted as attributes of the X/HTML tags, the semantic tags can then contain values that link specific elements of particular metadata vocabularies to the tagged text. This method has the advantage of tagging the information right in the text where it occurs. The example below from O'Donnell (2006, "Example" section) shows how Dublin Core might be used within the body of an XHTML page.

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">

<head profile="http://purl.org/NET/erdf/profile" >

<link rel="schema.dc" href="http://purl.org/dc/elements/1.1/" />

</head>

<body>

<h1 class="dc-title">Naked Metadata</h1>

<h2 class="dc-creator">Jonathan O'Donnell</h2>

<p class="dc-rights">http://purl.nla.gov.au/net/jod/tutorial/naked-metadata.html ©
Jonathan O'Donnell <span class="dc-date">23 October 2005<span></p>

</body>

</html>
```

A variety of tools exist to either generate or use semantically tagged data. DC-dot generates some semantic markup in X/HTML, RDF or XML for an existing web page without any (n.d.). Other projects, like GEM and PHIN, have implemented search interfaces for their own vocabularies. Extensions for Firefox, like Operator (n.d.) and Piggy Bank (2008), can extract data from web pages tagged with Microformats markup. Tools run the gamut of sophistication from simple scripts like eRDF detector (Alexander, 2007) to full-fledged data processors like ARC (2007).

However, despite the seeming multitude of tools available, this is where the infancy of the semantic web technology is most obvious. No one standard dominates the industry. There are still a variety of ways to have data semantically encoded without any correspondence, necessarily, between them. Many of the tools available are aimed at programmers and not end-users who just want to code a web page to provide access for other users, not write an entire customized suite of scripts to create and process the metadata generated. Development has begun, but has a long way to go before the semantic web will be ready for the mainstream.

How Does this Relate Specifically to Libraries?

The obvious question, for me, is how does the semantic web relate to libraries? It turns out very closely. First, the issue nearest to my heart: the semantic web has the potential to revolutionize cataloging. Cataloging could take place in software built using the same type of structure that would be used to build metadata into and extract it from web pages. Then cataloging metadata would be interoperable with all other RDF-compliant metadata on the web and, furthermore, library resources would be findable and usable in the same way as all other RDF-compliant web resources. The headaches brought on by trying to aggregate data from different, proprietary systems could become a distant memory. Different software packages to decode the individual markup schemes might not be necessary. (And there will be flowers and unicorns. People make the same claims about

open source software and that has yet to make major inroads into the ILS market. However, hope springs eternal.)

It could, in fact, come to mean that instead of using text strings to assign meaning to things like the author's name, cataloging would be a process of identifying the URI where an author's name is defined and including that URI in the cataloging record in place of a text string. (An example from the Library of Congress Subject Headings would be the URI "<http://id.loc.gov/authorities/sh98007973#concept>" in place of the text "Smith-Purcell effect.") The actual text of the author's name would not be retrieved until a user calls up that record for viewing. The advantage of this scenario is that name authority information only has to be updated once, in one place.

In this bewitching vision, we would share in the creation of Uniform Resource Identifiers (URIs) for works, expressions, manifestations, persons, corporate bodies, places, subjects, and so on. At the URI would be found all of the data about that entity, including the preferred name and the variant names. If any of that data needed to be changed, it would be changed only once, and the change would be immediately accessible to all users, libraries, and library staff by means of links (Yee, 2009, p. 55).

Tillett (2003) envisions an international virtual name authority file. "One proposal is to link the personal name authority file of the Library of Congress and that of the Deutsche Bibliothek (DDB) in Germany" (p. 115). The authority data linkages would be harvested and stored at a central location but the "day-to-day record maintenance activities continue to be managed as they are now by the national bibliographic agency (or regional authority)" (p. 116).

Another way that the semantic web fits in with the mission of libraries is that the core of the semantic web is built on controlled vocabularies, something librarians have been working with forever. Library work is full of pre-existing thesauri and controlled vocabularies that need only to be fitted into the appropriate structure for use as part of the semantic web.

The Semantic Web communities and library communities have both been working toward the same set of goals: naming concepts, naming entities, and bringing different forms of those names together. The tools and vocabularies developed in libraries, particularly those developed by the Library of Congress, are sophisticated and advanced. When translated into Semantic Web technologies they will help to realize Berners-Lee's vision (Harper & Tillett, p. 48).

The more connections that are made between things described in various places, the more value they have collectively. Tillett supports that idea. "We already have controlled vocabularies in our various authority files. Those could be linked with other controlled vocabularies of abstracting and indexing services, of biographical dictionaries, of telephone directories, and many other reference tools and resources to help users navigate" (p. 116-17).

Finally, Harper & Tillett quote Miller in stating that libraries have a role in developing the "layer of trust" for the semantic web. They are not clear about what that role might be, simply stating that "libraries have long standing trusted position that is applicable on the Web [*sic*]" (p. 50). Hillman (2008), on the other hand, provides a hint to how libraries might help build the trust layer:

A description consisting of aggregated sourced statements is susceptible to a variety of processes designed to provide downstream users with configurable descriptions based on their needs and capabilities. Over time, multiple statements can be rated using various criteria, and only the "best" used for exposure to downstream users who would rather not do the rating themselves. (p. 75)

By rating the descriptive statements that apply to any given resource, the library sets itself up as an authority on what information is "good" or "bad."

Issues and Problems

Unfortunately, it is never as simple as just issuing some standards and letting the developers take over. First of all, metadata has traditionally been centrally created by trained staff and, additionally, created after the fact. So two immediate challenges present themselves: getting the people who create the resources to also create standardized, coherent metadata to go with those resources and to get those creators to create that metadata at the same time as they create the resources. Greenberg, Sutton & Campbell (2003) state that "the glories of the Semantic Web will ultimately depend on tools that will enable authors to create with very little effort RDF annotations and other useful semantic metadata on their Web pages" (p. 18). At the current time, the tools allowing the web page developer to easily mark up applicable metadata are still being developed. As Coyle (2008) notes, the semantic web is still stuck in "engineer mode."

The documents on the [W3C] Semantic Web site develop concepts, set rules, and illustrate code. But even the most basic explanatory document, the RDF Primer, lacks examples of what services could be provided and how it might look to a user of the Semantic Web. (p. 264)

A second issue with the model of the semantic web is the question of bandwidth and accessibility. If there is no such thing, for example, as an "in-house" catalog record, what level of bandwidth is required for a system to retrieve and/or process each piece of data that must be retrieved from some other source on the Internet? A library, again for example, does not have one person calling up one record at a time; it would be operating on the scale of dozens of people calling up dozens of records each. How would this impact a place that can barely even afford its Internet connection with the existing bandwidth load? D'Arcus as quoted by Yee seems to imply that the way to get around this is to not retrieve the information in real-time. According to D'Arcus it can be retrieved in off-hours and stored locally for retrieval by users (Yee, p. 65). However, this solution creates its own problems, primarily that the data is now stored in two separate places and if changed in one, also needs to be changed also in the other.

Related to this is the very ephemeral existence of many sites on the Internet. Presumably, anyone who sets themselves up as an authority, i.e. creates a vocabulary, profile, registry etc., can somehow guarantee their continuing presence for some time at least, but unlike with traditional print authorities, once the website goes down, the authority information will be completely gone as well. Abrams (2005) also worries about fragmentation of authority information. In the context of a discussion about registries for digital file formats, he states:

It appears likely that many similar format registries may be developed or at least deployed at institutions around the world. This could result in an undesirable fragmentation

of important format representation information that would unnecessarily complicate the process of discovery of relevant data. (p. 132-3)

Another issue is that there is no one technology or metadata standard that fits everything. Unlike with MARC, there is no one ubiquitous standard that everyone will use. Some standards are not even developed yet, while others are still evolving. Hillman notes that even within communities, agreement on just a representative vocabulary can be difficult.

For the most part communities using metadata are still floundering in their attempts to figure out where the best balance between "rich and comprehensive" and "efficient and functional" can be defined. Part of the challenge is that few communities of practice have been able to define their needs as a community and take the next steps to implement services that support their goals. (p. 68)

Finally, the model for metadata exchange outside of MARC, as exemplified by Dublin Core, currently, is to develop detailed and careful descriptions for in-house use and make only simplified data available for exchange. Dublin Core is designed with core elements and qualifiers. Institutions can use the qualifiers for their own cataloging, but anything designed to be transferable has to make sense using only the core elements. The DCMI page *Using Dublin Core* states that the "element value (minus the qualifier) must continue to be generally correct and useful for discovery" (section 1.2). Campbell & Fast note that:

> Such an approach is highly useful to the development of interoperability standards…. However, while this approach goes a long way towards ensuring smooth and effective delivery of information across the Web, it does not necessarily allow libraries to exploit and contribute to the emerging Semantic Web in a full and exciting way. (p. 383)

Campbell and Fast continue:

> If we look at the Semantic Web merely as a cheap mechanism for exchanging metadata between approved metadata providers, we are shutting ourselves out from its potential richness. A Web that boasts a wealth of information, semantically coded and with a global addressing system, could be a source of cataloguing data in and of itself. (p. 386)

Conclusion: Challenges for the Future

The semantic web is coming. There is a multitude of developers all over the world working busily to ensure that. The first challenge is to get the tools into place for users of all levels. No one but the geekily inclined is going to be willing to invest the time and effort necessary to code semantically tagged web pages from scratch. The second challenge is getting semantic data into the everyday workflow of librarians. Semantic linkages with web content will enrich current content in ways yet unknowable. It is my intent to explore a small corner of the semantic web and contribute to the creation of linkages with a project to add metadata to pages on the site of cbsr.ucr.edu.

While decentralizing cataloging by pushing facets of it out to the content-creators is not going to actually put catalogers out of business, as so many fear, I do believe it will change how the job is done. Catalogers will have to become metadata experts, not just

MARC experts. Knowledge of RDF, XML and a variety of other metadata vocabularies (and the tools using them) will be necessary. Professional cataloging might be more a job of aggregating and improving harvested or contributed metadata, rather than developing new metadata, like MARC records, for resources. Hillman proposes that metadata evaluation will also be a large part of the future cataloger's job description.

> Increasingly … they may find themselves managing data from multiple sources, aggregating that metadata to serve a particular purpose, often not the one for which the metadata was originally created. Because most metadata available for aggregation, whether within an institution or via harvest, was created in a context most likely rife with assumptions that it would be used narrowly and only in a specific context, any aggregation project automatically involves some confrontation with metadata quality issues. (p. 66)

Cataloging has already entered a transitional phase, with more and more positions being advertised for metadata librarians rather than traditional cataloging librarians. It will be a challenge going forward for librarians and the content-creators to navigate this new re-imagined world. Cataloging is dead. Long live cataloging.

References

About microformats. (n.d.). Retrieved October 4, 2009 from http://microformats.org/about

Abrams, S. L. (2005). Establishing a Global Digital Format Registry. *Library Trends, 54* (1), 125-143.

Alexander, K. (2007). eRDF detector. *Userscripts.org.* Retrieved October 17, 2009 from http://userscripts.org/scripts/show/8260

Angjeli, A., Isaac, A., Cloarec, T., Martin, F., Meji, L. van der, Matthezing, H., et. al. (2009). Semantic web and vocabulary interoperability: an experiment with illumination collections. *ICBC, 38* (2), 25-29.

ARC: RDF classes for PHP. (2007). Retrieved October 17, 2009 from http://arc.semsol.org/home

Arch-int, N. & Sophatsathit, P. (2003). A semantic information gathering approach for heterogeneous information sources on WWW. *Journal of Information Science, 29* (5), 357-374.

Campbell, D. G. & Fast, K. V. (2004). Academic libraries and the semantic web: What the future may hold for research-supporting library catalogues. *The Journal of Academic Librarianship 30* (5), 382-390.

Centers for Disease Control and Prevention. (2005). *Public Health Information Network vocabulary metadata standards. Version 1.2. 08/08/2005.* Retrieved October 11, 2009 from http://www.cdc.gov/phin/library/documents/pdf/PHIN%20Vocabulary%20Metadata%20V1.2.pdf

Chavarriaga, E. & Macias, J. A. (2009). A model-driven approach to building modern semantic web-based user interfaces. *Advances in Engineering Software, 40,* 1329-1334.

Coyle, K. (2008). Meaning, Technology, and the Semantic Web. *The Journal of Academic Librarianship, 34* (3), 263-4.

Damiani E. & Fugazza, C. (2007). Toward semantics-aware management of intellectual property rights. *Online Information Review, 31* (1), 59-72.

DC-dot: Dublin Core metadata editor. (n.d.). Retrieved October 17, 2009 from http://www.ukoln.ac.uk/cgi-bin/dcdot.pl?n=0&guesspublisher=yes

DCMI/RDA Task Group. (2008). *DCMI/RDA Task Group Wiki.* Retrieved October 10, 2009 from http://dublincore.org/dcmirdataskgroup/

Dublin Core Metadata Initiative. (2005) Using Dublin Core. Retrieved October 17, 2009 from http://dublincore.org/documents/usageguide/

Dublin Core Metadata Initiative. (2008). *The Dublin Core metadata registry: Promoting the discovery and reuse of metadata.* Retrieved October 10, 2009 from http://dcmi.kc.tsukuba.ac.jp/dcregistry/

Gateway to Educational Materials information. (2009). Retrieved October 17, 2009 from http://www.thegateway.org/about

Global Digital Format Registry. (n.d.). Retrieved October 17, 2009 from http://www.gdfr.info/

Greenberg, J., Sutton, S. & Campbell, D. G. (2003). Metadata: A fundamental component of the Semantic Web. *Bulletin of the American Society for Information Science and Technology, 29* (4), 16-18.

Guzmán Luna, J., Torres Pardo, D. & López García, A. N. (2006). Desarrollo de una ontología en el contexto de la web semántica a partir de un tesauro documental tradicional. *Revista Interamericana de Bibliotecología 29* (2), 79-95

Harper, C. & Tillett, B. (2007). Library of Congress controlled vocabularies and their application to the semantic web. *Cataloging & Classification Quarterly, 43* (3), 47-68.

Heery, R. & Wagner, H. (2002). A metadata registry for the semantic web. *D-Lib Magazine, 8* (5). Retrieved September 20, 2009 from http://www.dlib.org/dlib/may02/wagner/05wagner.html

Hildebrand, M., Ossenbruggen, J. van, Hardman, L. & Jacobs, G. (2009). Supporting subject matter annotation using heterogeneous thesauri: A user study in Web data reuse. *International Journal of Human-Computer Studies, 67,* 887-902.

Hillman, D. (2008). Metadata Quality: From Evaluation to Augmentation. *Cataloging & Classification Quarterly, 46* (1), 65-80.

International Federation of Library Associations and Institutions (2008). *Declaring FRBR entities and relationships in RDF.* Retrieved October 10, 2009 from http://www.ifla.org/files/cataloguing/frbrrg/namespace-report.pdf

JISC IE Metadata Schema Registry. (2009). Retrieved October 17, 2009 from http://www.ukoln.ac.uk/projects/iemsr/

Library of Congress. (n.d.). About [authorities & vocabularies]. Retrieved October 17, 2009 from http://id.loc.gov/authorities/about.html

Microformat. (2009). *Wikipedia.* Retrieved October 4, 2009 from http://en.wikipedia.org/wiki/Microformats

The National Archives. (n.d.). *The technical registry: PRONOM.* Retrieved October 10, 2009 from http://www.nationalarchives.gov.uk/PRONOM/Default.aspx

National Science Digital Library. (2009). *NSDL registry: Supporting metadata interoperability.* Retrieved October 10, 2009 from http://metadataregistry.org/

O'Donnell, J. (2006). *Naked Metadata.* Retrieved October 17, 2009 from http://jod.id.au/tutorial/naked-metadata.html

Operator. (n.d.). *Mike's Musings.* Retrieved October 17, 2009 from http://www.kaply.com/weblog/operator/

Piggy Bank. (2008). Retrieved October 17, 2009 from http://simile.mit.edu/wiki/Piggy_Bank

Talantikite, H. N., Aissani, D. & Boudjlida, N. (2008). Semantic annotations for web services discovery and comparison. *Computer Standards & Interfaces, 31,* 1108-1117.

Talis. (2006). *Rdf in html.* Retrieved October 12, 2009 from http://research.talis.com/2005/erdf/wiki/Main/RdfInHtml

Tillett, B. (2003). AACR2 and Metadata: Library opportunities in the global semantic web. *Cataloging & Classification Quarterly, 36* (3), 101-119.

Tonkin, E. & Strelnikov, A. (2009). Spinning a semantic web for metadata: Developments in the IEMSR. *Ariadne, 59.* Retrieved October 17, 2009 from http://www.ariadne.ac.uk/issue59/tonkin-strelnikov/

Uddin, M. N. & Janecek, P. (2006). Faceted classification in web information architecture: A framework for using semantic web tools. *The Electronic Library 25* (2), 219-233.

Unified Digital Format Registry (UDFR). (2009). Retrieved October 17, 2009 from http://www.udfr.org/

World Wide Web Consortium (2004a). *RDF primer.* Retrieved October 10, 2009 from http://www.w3.org/TR/rdf-primer/

World Wide Web Consortium (2004b). *RDF vocabulary description language 1.0: RDF Schema.* Retrieved October 4, 2009 from http://www.w3.org/TR/rdf-schema/

World Wide Web Consortium (2004c). *RDF/XML syntax specification (revised.* Retrieved October 17, 2009 from http://www.w3.org/TR/rdf-syntax-grammar/

World Wide Web Consortium (2007). *Gleaning resource descriptions from dialects of languages (GRDDL).* Retrieved October 4, 2009 from http://www.w3.org/TR/grddl/

World Wide Web Consortium (2008a). *RDFa primer: Bridging the human and data webs.* Retrieved October 4, 2009 from http://www.w3.org/TR/xhtml-rdfa-primer/

World Wide Web Consortium (2009b). *SKOS simple knowledge organization system primer.* Retrieved October 10, 2009 from http://www.w3.org/TR/skos-primer/

World Wide Web Consortium (2008c). *SPARQL query language for RDF.* Retrieved October 10, 2009 from http://www.w3.org/TR/rdf-sparql-query/

World Wide Web Consortium (2009a). *Protocol for web description resources (POWDER): W3C working group note 1 September 2009: Primer.* Retrieved October 4, 2009 from http://www.w3.org/TR/powder-primer/

World Wide Web Consortium (2009b). *Semantic web: W3C semantic web frequently asked questions.* Retrieved October 4, 2009 from http://www.w3.org/RDF/FAQ

World Wide Web Consortium (2009c). *Semantic web: Web ontology language (OWL).* Retrieved October 4, 2009 from http://www.w3.org/2004/OWL/

Yee, M. (2009). Can Bibliographic Data be Put Directly onto the Semantic Web? *Information Technology and Libraries, 28* (2), 55-80.

Additional Resources

Adams, K. (2002). The semantic web: Differentiating between taxonomies and ontologies. *Online, 26* (4), 20-23.

Berners-Lee, T. (2009 May) Linked Data [video file]. Presentation at TED2009: "The Great Unveiling" Conference, Long Beach, California, USA. Retrieved October 4, 2009 from http://www.ted.com/index.php/talks/tim_berners_lee_on_the_next_web.html

Bradley, F. (2009). Discovering linked data. *Library Journal, 134* (7), 48-50.

Chudnov, D. (2009). The geography of linked data and ready reference. *Computers in Libraries, 29* (4), 26-28.

Coombs, K. (2009). Microformats: Context inline. *Library Journal, 134* (7), 64.

Coyle, K. (2009). Making connections. *Library Journal, 134* (7), 44-47.

DOAP: Description of a Project (n.d.). Retrieved October 17, 2009 from http://trac.usefulinc.com/doap

Dublin Core Metadate Initiative. (2009). *DCMI metadata terms.* Retrieved October 10, 2009 from http://www.dublincore.org/documents/dcmi-terms/

Dumbill, E. (n.d.). *DOAP.* Retrieved October 10, 2009 from http://trac.usefulinc.com/doap

EAD: Encoded Archival Description. (2009). Retrieved October 10, 2009 from http://www.loc.gov/ead/

Embedded RDF (2009). Wikipedia. Retrieved October 12, 2009 from http://en.wikipedia.org/wiki/Embedded_RDF

Feigenbaum, L. (2009 May). *The 2009 semantic web landscape* [slideshow]. Presentation at the PRISM Forum SIG Meeting, Luzern, Switzerland. Retrieved October 4, 2009 from http://www.slideshare.net/LeeFeigenbaum/semantic-web-landscape-2009

Fichter, D. & Wisniewski, J. (2008). Microformats and the search for meaning. *Online, 32* (4), 55-57.

The Friend of a Friend (FOAF) project. (n.d.). Retrieved October 10, 2009 from http://www.foaf-project.org/

Gradmann, S. (2005). Rdfs:frbr– Towards an implementation model for library catalogs using semantic web technology. *Cataloging and Classification Quarterly, 39* (3), 63-75.

Herman, I. (2009a June). *Introduction to the Semantic Web (tutorial)* [slideshow]. Presentation at the 2009 Semantic Technology Conference, San Jose, CA, USA. Retrieved October 4, 2009 from http://www.w3.org/2009/Talks/0615-SanJose-tutorial-IH/Slides.pdf

Herman, I. (2009b June). *What is new in W3C land?* [slideshow]. Presentation at the 2009 Semantic Technology Conference, San Jose, CA, USA. Retrieved October 4, 2009 from http://www.w3.org/2009/Talks/0615-SanJose-talk-IH/Slides.pdf

Hodge, G. (2000). *Systems of knowledge organization for digital libraries: Beyond traditional authority files.* Council on Library and Information Resources. Retrieved October 4, 2009 from http://www.clir.org/pubs/reports/pub91/contents.html

Kolbitsch, J. & Krottmaier, H. (2006). *The Use of HTML-Encoded Dublin Core in Academic and Educational Settings.* Retrieved October 17, 2009 from http://www.kolbitsch.org/research/papers/2006-Dublin_Core_Analysis.pdf

Krishnamurthy, M. (2006). Library portals: Ontological representation of knowledge on the web. *Information Studies, 12* (2), 75-84.

Miller, R. (2008). Content delivery rides the semantic web. *EContent, 31* (8), 26-30.

Msporny. (2008). *RDFa basics* [Video file]. Retrieved October 4, 2009 from http://www.youtube.com/watch?v=ldl0m-5zLz4&NR=1

OAIster …find the pearls. (2009). Retrieved October 17, 2009 from http://www.oaister.org/

RDFa. (2009). *Wikipedia.* Retrieved October 10, 2009 from http://en.wikipedia.org/wiki/Rdfa

Rogers, G. P. (2007). Roles for semantic technologies and tools in libraries. *Cataloging & Classification Quarterly, 43* (3), 105-125.

Smith, G. (2009). Web 3.0: 'Vague but exciting'. *AdweekMedia, 50* (24), 19.

Social network service. (2009). *Wikipedia.* Retrieved October 17, 2009 from http://en.wikipedia.org/wiki/Social_network_service

Sutton, S. (2008). Metadata quality, utility and the semantic web: The case of learning

resources and achievement standards. *Cataloging & Classification Quarterly, 46* (1), 81-107.

TEI: Text Encoding Initiative. (2009). Retrieved October 10, 2009 from http://www.tei-c.org/index.xml

Wiki. (2009). *Wikipedia.* Retrieved October 17, 2009 from http://en.wikipedia.org/wiki/Wiki

World Wide Web Consortium (2005). *Quick guide to publishing a thesaurus on the semantic web.* Retrieved October 4, 2009 from http://www.w3.org/TR/2005/WD-swbp-thesaurus-pubguide-20050517/

World Wide Web Consortium (2009a). *Protocol for web description resources (POWDER) working group.* Retrieved October 4, 2009 from http://www.w3.org/2007/powder/

World Wide Web Consortium (2009b). *Semantic web: W3C semantic web activity.* Retrieved October 4, 2009 from http://www.w3.org/2001/sw/

World Wide Web Consortium (2009c). *Semantic web: W3C semantic web activity publications.* Retrieved October 4, 2009 from http://www.w3.org/2001/sw/Specs.html

Folksonomies for Digital Resources

Jill Bauhs Jensen

Jill Bauhs Jensen can be reached at: jillieb@hotmail.com

Introduction

With the advent of the second wave of Internet functionality (Web 2.0) came the ability for users to apply "tags" or descriptive words to digital resources, categorizing them and allowing for easier retrieval by both the tagger and other users. Considering the vastness of the Internet, does user tagging hold the key to cataloging the Web? "In order to manage the Web's massive amount of data, the process of social networking, through tagging, becomes an appealing option, especially since this tagged information is practically labeled by the user and can be shared" (Snipes, 2007).

Folksonomy Defined

"The collection of user-assigned tags is referred to commonly as a folksonomy" (Spiteri, 2007). The word "folksonomy" is rooted in the word taxonomy, but rather than

describing a hierarchical organizational system, it denotes a flat system "of the people", where all terms are equal. Creator Thomas Vander Wal elaborates:

Folksonomy is the result of personal free tagging of information and objects (anything with a URL) for one's own retrieval. The tagging is done in a social environment (shared and open to others). The act of tagging is done by the person consuming the information.

The value in this external tagging is derived from people using their own vocabulary and adding explicit meaning, which may come from inferred understanding of the information/object as well as. The people are not so much categorizing as providing a means to connect items and to provide their meaning in their own understanding.

Tags are often aggregated to create searchable metadata, creating broad, general categories. Suster (2006) defines tagging as "a democratic and distributed classification method." User-generated metadata stands in contrast to traditional cataloging, which places content into pre-defined categories and sub-categories (Shirky).

Occasionally, sites that permit social tagging include tag clouds, or visual representation of the most used tags. The larger the font in the tag cloud, the more prominent that description is within the metadata (Snuderl, 2008; Steele, 2009).

Tags generally fall into one of seven types: descriptive (subject of the resource), type ("video", "blog", "image"), ownership, opinion of the resource, self-reference ("myarticle"), organizational (e.g., course number) and lastly playful ("squaredcircle") (Smith, 2008)

Cataloging Digital Assets

Today, our world abounds with vast digital assets in the form of written words, images, sound and video, and much of this information remains uncatalogued by information professionals. In addition, conventional library organizational structures are stilted, and not conducive to natural language searches users are accustomed to, thanks to search engines like Yahoo and Google (Buckland). Can user-generated metadata bridge the divide between traditional classification and user expectations that search language be both flexible and current? If so, can the quality of tagging be enhanced to aid in both precision and recall when searching?

Literature Review

Undoubtedly because the field is new, the literature on folksonomies focuses primarily on the theoretical advantages and disadvantages of user-generated metadata. In addition, several groups have undertaken examinations of tagging on commercial sites, and integration of folksonomies with traditional cataloging in library settings. Finally, several writers debate whether user tags are worth investing in at all.

Advantages and Disadvantages of User-Generated Metadata

Much of the literature on user-generated metadata to date focuses on the benefits and pitfalls of folksonomies. Numerous advantages to using folksonomies are detailed. High on the list is inclusiveness:

"Social tagging's great strength lies in its openness; the terms are devised and implemented by actual users, so by definition, they reflect the language of those users (McElfresh, 2008).

Spiteri (2007) agrees. "They reflect the vocabulary of the users, regardless of viewpoint, background, bias, and so forth". Because they are in the vernacular, and from the point of view of the lay person, user tags can be more comprehensive than cataloging created by subject experts:

User tags could provide additional points of view to that in existing museums records. Museum documentation is known to address works of art from a perspective different than that of the public. Within the context of art museums, user contributed tags might help reflect the breadth of approaches to works of art, and offer access to alternative points of view. (Trant, 2008)

Kroski (2007) adds that "because folksonomies include alternative views together with popular ones, they present a unique opportunity to discover … interests of the minority that lie at the 'tail' end of a power law, or statistical distribution." West (2007) believes that this leads to " more access points and richer metadata, which results in the materials being more findable."

Hand-in-hand with inclusiveness is the ability of folksonomies to be updated quickly. McElfresh (2008) deems them "nimbler and more flexible than controlled vocabularies" and later adds:

"Changes to controlled vocabularies like Library of Congress subject headings happen at a pace constrained by the machinery of the bureaucracies that create the classification system (McElfresh, 2008).

Kroski sees similar benefits. Folksonomies are current and flexible, and there is no need to predict categories in advance. They offer the potential to discover "unknown and unexpected resources," leading to exploration. Object can be tagged in multiple categories, offering multifaceted richness. Tags follow "desire lines." "It's not about the right or the wrong way to categorize something and it's not about accuracy or authority, it's about remembering".

In addition, folksonomies are simple (McElfresh), low-cost (Kroski, Furner), easily used (Kroski), empowering (Furner), share the workload (McElfresh) and create a spirit of sharing within the community (Kroski, West).

One final benefit is overlooked by most of the literature: tags also link directly to the tagged resource. "Tagging is an online, hyperlinked activity. When an item such as a bookmark, a picture, or even a person's profile has a tag added to it, the tag becomes a clickable link to more items associated with that tag at either a personal or a system-wide level" (West). While this may seem simple, even obvious, this is the equivalent to the card catalog automatically retrieving the book from the shelf for the user, a huge benefit.

Folksonomies are not without their disadvantages. One of the most interesting criticisms is that tags, by their very democratic nature, can be contradictory.

If I tag an article with the subject "white horse" and you tag it "black horse", that is all right since both can coexist in a folksonomy classification scheme. The problem with relativism is the question: "relative to what?" Each Internet user is bringing to bear on the item a different linguistic and cultural background. Although this is an inherent strength of folksonomies (since it recognizes many valuable individual perspectives), it can also lead to the existence of contraries. (Peterson, 2006)

Spiteri (2007), meanwhile, sees problems with synonymy. A word like "port" has multiple meanings, ranging from a type of wine to definitions relating to ships. The same concept may have different spellings (West). Expanding on this idea, Steele, 2009) writes:

"A similar problem with tagging is synonymy. While this can be something as simple as using the tag "TV" instead of "television," many terms have even more synonyms, such as "pop", "soda," "coke", "soft drink", or even "soda pop". A controlled vocabulary handles this issue again with an authority file. (Steele, 2009)

Conversely, "polysemy is another problem with tagging related to the tagger's vocabulary selection. In this case, however, the user may select a word that has more than one similar meaning" (Steele, 2009). Spiteri uses the example of Apple Macintosh computers. Users may select tags like "mac," "Macintosh" and "apple" to describe the same device.

Both Spiteri and Steele point out that plurality is another potential pitfall. "Tagging would rely on the user to search both the singular and plural forms, since the original tagger would be likely to enter the tag in only one of the forms" (Steele, 2009).

Kroski (2007) adds that folksonomies lack hierarchy, and taxonomies "provide a deeper, more robust classification of entities. Such systems allow users a finer granularity in searching for resources". Complete recall is also affected. "Because of the lack of synonym control, a folksonomy search will not effect a complete results list because of the use of similar tags" (Kroski). "With tagging, it is up to the user to tag for both the broader and narrower terms if the resources will be retrieved" (Steele, 2009).

Malicious activity is another concern (Snipes, 2007). "A user can cause harm by tagging resources with inappropriate terms" (Steele, 2009). "With social tagging, the very openness that empowers users (or at the very least, draws people in) also leaves the tagging system - and any classification systems built upon it - exposed to vandalism or other abuse" (McElfresh, 2008).

Finally, Snipes (2007) is concerned that the tags be credible (in other words, accurate) and consistent, and respect the privacy of the user.

Despite the limitations of folksonomies, the potential exists for far-reaching effects of user tagging of electronic resources. As Peterson states, "Applying folksonomy tags has the potential to be very popular" (2008, p.4).

Commercial Sites Using Folksonomies

Several commercial sites that have implemented user tagging are frequently mentioned in the literature. They include del.icio.us, flickr.com, and librarything.com. West describes them plainly:

del.icio.us - social bookmarking where users of the site globally share bookmarks tagged with words like "todo" or "tools" or "coupons" to help them find and remember their bookmarks.

flickr.com - photo sharing where people add tags to their images as a form of textual metadata to both help them find their own photographs as well as locate similarly tagged photographs by others.

librarything.com - online personal library where people can add tags like "toread" or "thriller" or "topten" to the books they enter into their online libraries and see who else has books with the same or similar tags.

According to McGregor (2006), "del.icio.us is arguably the most developed and possibly the most collaborative." When the user is tagging a new URL, the site provides a list of "popular" tags that were previously used to describe that resource. "These common tags can then be used in a subsequent user search strategy" (McGregor). Detailing Golder and Huberman's study of del.icio.us tagging:

They found that the users of collaborative tagging systems exhibited much variety in the sets of tags they employ. The frequency of tag use and what the tags themselves described was also found to vary greatly between users. However, the data also suggested that there existed some measure of regularity in the tags being assigned by users.

Spiteri (2007) used the daily tag logs from d el.icio.us, among other sites, comparing the tags with "the National Information Standards Organization (NISO) guidelines for the construction of controlled vocabularies." She found that the tags and guidelines had a close correspondence in some areas (concepts, single terms, spelling) but that count nouns and ambiguous tags still presented problems. Her conclusion was that providing del.icio.us users with additional guidelines and links to reference sources could mitigate some problems. "Folksonomies could serve as a powerful, flexible tool for increasing the user-friendliness and interactivity of public library catalogs."

Flickr.com allows users to form collections of images, and then tag those images for either public or private viewing. According to Dye (2006):

On Flickr, the focus is less on how to promote something within a community and more on how to increase the findability of personal content. … Although the burden of creating metadata rests mostly with the person who posts the content, social groups can use Flickr to create group tags to collect all members' photos tagged with a particular keyword together—something that Flickr calls *tagography.*

The LibraryThing website describes itself as "an online service to help people catalog their books easily" (LibraryThing, 2009). It incorporates data from Amazon, the Library of Congress and other world libraries in assisting users in the cataloging process. Rolla (2009) conducted a study comparing " LibraryThing tags for a group of books and the library-supplied subject headings for the same books." Not surprisingly, he found that professional catalogers and LibraryThing users described resources very differently. Rather than viewing

this as a problem, he believes that "Because of these differences, user tags can enhance subject access to library materials, but they cannot entirely replace controlled vocabularies such as the Library of Congress subject headings." Yet, he concludes that " adding user tags to library catalogs could help improve subject access to collections."

Incorporating Folksonomies in Library Settings

Of course, library professionals are most interested in the possible inclusion of user-generated metadata within more traditional library catalogs. "We could harness folksonomies to build 'hybrid catalogs,' strengthening the catalog services we provide to our patrons" (McElfresh, 2008).

Spiteri (2006) outlines three ways that libraries can implement user tagging. First, "allow users of public library catalogues to create and organize their own personal information space in the catalogue" (p. 76). Adding their own notes, and references to outside links, enriches the user's personal catalog.

Second, "Allow users to supplement the existing controlled vocabulary in the catalogue … with their own metatags" (p. 76). Doing so allows for more natural (and current) language. And, third, "Folksonomies could be used to foster online communities of interest … public tags can be viewed by other catalogue users with similar interests; this use of folksonomies could facilitate the sharing and exchange of information" (pp. 76-77).

Kroski (2008) contributes additional means of using folksonomies. "Libraries are making use of social cataloging applications as tools to cataloging new titles in a browsable, interactive, user-focused community" (p. 81). The Danbury Library in Connecticut has integrated LibraryThing into its OPAC (Online Public Access Catalog), allowing users to tag and recommend books.

By implementing this service, the Danbury Library not only had channeled additional functionality for its library patrons, but incorporated a massive resource of focused user-generated content. This customizable service enabled them to enhance their user experience while maintaining the previous functionality of their OPAC.

Mendes, Quinonez-Skinner and Skaggs (2009) studied LibraryThing for Libraries at California State University, Northridge (CSUN) and found:

For every new book a user discovers using LCSH headings they will discover four books using LTFL. This type of data captures sheer numbers, with the potential for increased resource discovery. It does not address relevancy of additional books for users search. … The addition of user-generated metadata to catalog records, however, does enhance resource discovery, for example, for those titles lacking subject headings (e.g. works of fiction). Tags facilitate the discovery of resources by genre. Tags reflect the natural language of users and as such provide new paths for resource discovery.

The University of Pennsylvania had developed a closed system called PennTags that allow users within the Penn community to tag "web sites, articles in the library's database, and records in both the video catalog and Franklin, the library's OPAC" (Steele, 2009). "Folksonomy terms exist side by side with the LCSH headings … although they are not yet prevalent in the catalog" (Peterson, 2008). Unfortunately, users are unable to search the OPAC by these tags (Steele).

Steele documents several other libraries that are using folksonomies within their collections. Ohio State University Libraries are using one tag, *leisurereading,* "to create a list of books the library owns that patrons can read for leisure. … Once the patron clicks on the individual book, they can see several other tags created by other LibraryThing members." He also mentions that Ann Arbor District Library (AADL) has developed SOPAC or "social online public access catalog.

Finally, Montana State University is using folksonomies to tag Electronic Theses and Dissertations (ETDs). Peterson (2009), an Associate Professor and Information Resources Specialist at the University writes that " patrons are using folksonomy tags, and the usage of the tags is increasing." But, "it appears that the uses of LCSH and folksonomy are quite different, and that these parallel modes of access should continue to maximize usability and ease of access to the database."

Professional Skepticism

One of the more subtle hurdles user-generated metadata may need to overcome is the skepticism of some information professionals. Snipes (2007) enumerates some expert concerns. Are the tags credible? Can the information be trusted? What about malicious postings? How do users safeguard their privacy? "At this time, our students still need the security and dependability of the traditional, controlled information seeking methods." Yet, at the same time she concedes, "Folksonomy is here to stay, and it shows promise of evolving into a vital structure of information retrieval for our students in the future."

Other professionals have gradually warmed to the benefits of user-generated metadata. In 2006, Peterson expressed her misgivings:

Folksonomy allows for disparate opinions and the display of multicultural views; however, in the networked world of information retrieval, a display of all views can also lead to a breakdown of the system. … Folksonomy is a scheme based on philosophical relativism, and therefore it will always include the failings of relativism. A traditional classification scheme will consistently provide better results to information seekers.

In 2008, her position softened slightly. "… Subject cataloging and user-generated tags will probably coexist." Yet, within a year she concluded, after studying their usage at the Montana State University libraries, that "P atrons are using folksonomy tags, and the usage of the tags is increasing. … Usage of the tags is evidence that permitting folksonomy tags in the ETD database has met patrons' needs. "

While Peterson was concerned about philosophical relativism, Shirky's point of view is the opposite. "It comes down ultimately to a question of philosophy. … [If] you believe that we make sense of the world, if we are, from a bunch of different points of view, applying some kind of sense to the world, then you don't privilege one top level of sense-making over the other." The collective wisdom of the users of the system has inherent benefit.

Macgregor and McCulloch (2006) similarly see benefit in incorporating user tags into existing cataloging:

Librarians and information professionals have lessons to learn from the interactive and social aspects exemplified by collaborative tagging systems, as well as their success in

engaging users with information management. The future coexistence of controlled vocabularies and collaborative tagging is predicted, with each appropriate for use within distinct information contexts: formal and informal.

Or, as West (2007) concludes, "Anything that can help our users find information should be a net gain to librarians."

Improving Tagging Quality

This project addresses the question of whether providing subject matter experts who are not information professionals will improve the quality of their tagging when provided simple instructions. In other words, does guidance increase the value of tags?

Methodology

The commercial photo-sharing site Flickr.com was chosen as the platform for this study because of its ease-of-use and simple tagging feature. Two private Flickr accounts were created to allow users to upload photos with the understanding that they would not be repurposed for commercial or other uses. (Image theft has recently been an issue for some participants.)

Several user communities of Portuguese Water Dog enthusiasts were asked to participate. This group was chosen for several reasons, including their subject matter expertise and their access to digital photographs taken in a variety of settings, from formal dog shows and trials, to casual at-home photography.

Participants' demographics are similar to purebred dog enthusiasts in general: a typical contributor is female, middle-aged, unmarried with no children living at home.

Participants were told that their photos would remain private, but that they could tag other photos within the Flicker account as well as their own. A sample photo with tags was uploaded to the account for those given tagging instructions, but it was not identified as such to the participants. However, each participant could view all photos that had been uploaded to the account and model their tagging on other photos if they liked.

Contributors were divided into two groups and asked to upload one or more photos and tag them. The first group, as a control, was given no instructions beyond the web address and login information for their Flickr account. The second group was also given login information, and provided with the following instructions:

Tagging is a way of marking photos, articles, blogs and books so they can easily be searched. It allows similar items to be grouped together (like books on vampires). Tagging also helps differentiate between similar objects (a picture of a red ball vs. a blue ball).

Please consider the following ideas when you tag your photos, but feel free to use whatever tags you like:

Add a descriptive title: *Dad Rowing on Castaic Lake*

Add tags that describe your picture(s). Helpful tags might include title, creator (photographer), date, location, descriptive words … whatever you think will help in locating the photos later.

Separate each tag with a space: *cameraphone urban moblog.* To join 2 words together in one tag, use double quotes: *"daily commute".* All your tags don't need to be entered at once. If you think of more, you can add more.

Several of the users were uncomfortable uploading their own photos, and so they submitted their photos via email, and the pictures were uploaded to the Flickr account for them.

(Photos may be viewed using the login information found in Appendix A.)

Results

Twelve photos were uploaded to the control group account in addition to the sample photo, and each photo was tagged an average of three times. In addition, six of the twelve were given a descriptive title, either on the participant's local computer or when the photo was uploaded; five pictures were given descriptions.

By contrast, twenty-nine photos were uploaded to the account used by participants given instructions. Again, each photo averaged three tags, although one image has ten tags and seven have none. The majority of photos (25) have a descriptive title, and fourteen have descriptions.

The following table details tags and descriptions that were suggested in the instructions:

	Control Group	Instructions Group
Number of photos	12	29
Number of tags	38	89
Average number of tags	3.16	3.06
Descriptive title	6 (50%)	25 (86%)
Description	5 (41%)	14 (48%)
Breed tag (either Portuguese Water Dog or PWD)	12 (100%)	13 (45%)
Call name	3 (8% of tags)	14 (15% of tags)
Registered name	0	0
Creator tag	0	2
Date tag	0 (12 had auto-created date tags generated by the camera)	0 (21 had auto-created date tags generated by the camera)
Location tag	0	1 (location was identified in the title on two additional images)

(A complete list of tags for each group, and corresponding tag clouds, can be found in Appendix B.)

Observations

At first blush, it appears that the quantity of tags was virtually the same for both groups. However, when photos without tags are eliminated, the instruction group averages four tags per photo, compared with the control group's three per image. (Images without tags were primarily those uploaded on behalf of a contributor.) In addition, the photo descriptions in the instructions group were often highly detailed, containing data that would be included in tags by information professionals. For example, photo *IMG_4459* does not have a descriptive title, but the description contains detailed information, including call names, breeds and MPH of the dogs:

Kibble Power - 265 DP with 20 PAW DRIVE !!! (recently clocked at 15+ MPH!) The dog driving Porties (Annie, Gladys, Chuckie, Jibby) and Qika the Bouvier de Flanders.

Overall, there was a higher level of participation by those in the instructions group. This may be attributed to the inclusion of a definition of tagging in the instructions. (Participants' computer skills varied from extremely familiar with uploading digital photos to a complete lack of comfort in this area. However, when approached about participating most contributors indicated that they did not have a clear understanding of tagging, so a description of tagging was added to the instructions.)

Whether the instructions led to a higher quality of tags, contributing to better precision and recall, is hard to determine. It is apparent that a much larger group of participants would be required to make this determination. But, the inclusion of instructions did lead to much more creative and varied tags. Rather than restricting the efforts of the contributors, including instructions appears to have allowed them more flexibility in their tagging. Delightful tags like " expression", "face", "kibblepower" and "watchyourdrinkaroundhim" add richness to the image descriptions.

One of the surprises of this study was how few of the participants included basic information about the dogs, including breed, kennel and call name (e.g. "Trixie"). In addition, the registered name of the dog, the definitive description of a purebred dog, was not used as a single tag. This was highly unexpected from a group of experts where this information begins even the most casual of conversations.

Conclusion

While this study of concerning improving the quality of user tagging is not definite, a few conclusions can be drawn. Users may respond more creatively in their tagging when they are given information explaining what tagging is and offering examples. Rather than feeling hemmed in by instructions, they appear to be more willing to expand their tagging horizons.

For many users, describing a resource is less compartmentalized than it might be for an information professional. Lay users treat title, description and tags as interchangeable, and search software (and perhaps the tagging mechanism itself) should accommodate this fluidity, even exploit it.

Finally, participants are more enthusiastic when offered some instruction, which in turns leads to higher levels of participation and creativity. Providing direction may be the

first step in harnessing user-generated tagging for supplemental cataloging of digital resources.

The frontier of user-generated metadata is only now being explored. But, the possibilities are intriguing. "The overall usefulness of folksonomies is not called into question; just how they can be refined without losing the openness that makes them so popular" (Peterson 2006).

Appendix A

Login Information

Photos may be viewed at www.flickr.com using the following login information.

Control Group:

Username pwdphotos1

Password trixie1

Instructions Group:

Username pwdphotos2

Password portie2

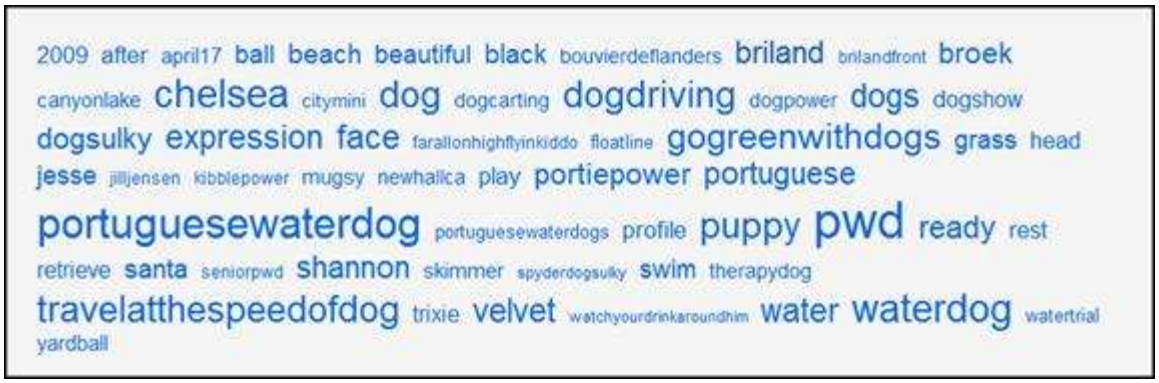
Appendix B

Tag Clouds and Tag Lists

Control Group:

Tag	Entered as	On
Agility	Agility	1 item
dallas	dallas	1 item
dog	dog	5 items
dogs	Dogs	1 item
lioncut	Lion cut	1 item
portuguese	portuguese, Portuguese	6 items
portuguesewaterdog	portuguese water dog Portuguese Water Dog	2 items
portuguesewaterdogs	Portuguese Water Dogs	3 items
puppy	puppy	2 items
pwd	pwd, PWD	6 items
pwdpuppy	PWD puppy	1 item
ru	ru	1 item
tessa	tessa	1 item
trixie	Trixie	1 item
water	water, Water	6 items
ziggylitter	ziggy litter	1 item

Instruction Group:



Tag	Entered as	On
2009	2009	1 item
after	After	1 item
17-Apr	17-Apr	1 item
ball	ball	1 item
beach	beach	1 item
beautiful	Beautiful	1 item
black	Black	1 item
bouvierdeflanders	Bouvier de Flanders	1 item
briland	Briland	2 items
brilandfront	Briland-front	1 item
broek	Broek	2 items
canyonlake	Canyon Lake	1 item
chelsea	Chelsea	4 items
citymini	City Mini	1 item
dog	Dog, dog	3 items
dogcarting	dog carting	1 item
dogdriving	dog driving, dogdriving	3 items
dogpower	dog power	1 item
dogs	dogs	2 items
dogshow	dog show	1 item
dogsulky	dog sulky	2 items
expression	expression	2 items
face	face	2 items
farallonhighflyinkiddo	Farallon High Flyin' Kiddo	1 item
floatline	float line	1 item
gogreenwithdogs	Go Green with dogs gogreenwithdogs	3 items
grass	grass	1 item
head	head	1 item
jesse	Jesse	1 item
jilljensen	Jill Jensen	1 item
kibblepower	Kibble Power	1 item
mugsy	Mugsy	1 item
newhallca	Newhall, CA	1 item
play	play!!!	1 item
portiepower	portie power, portiepower	2 items
portuguese	Portuguese	2 items

portuguesewaterdog	Portuguese Water Dog portuguesewaterdog	5 items
portuguesewaterdogs	Portuguese Water Dogs	1 item
profile	profile	1 item
puppy	puppy	3 items
pwd	PWD, pwd	8 items
ready	ready	2 items
rest	rest.	1 item
retrieve	retrieve	1 item
santa	Santa	1 item
seniorpwd	senior pwd	1 item
shannon	Shannon	2 items
skimmer	Skimmer	1 item
spyderdogsulky	Spyder dog sulky	1 item
swim	Swim	1 item
therapydog	therapy dog	1 item
travelatthespeedofdog	Travel at the Speed of Dog travelatthespeedofdog	3 items
trixie	Trixie	1 item
velvet	Velvet	2 items
watchyourdrinkaroundhim	watch your drink around him	1 item
water	Water	2 items
waterdog	Water Dog, waterdog	4 items
watertrial	water trial	1 item
yardball	yard-ball	1 item

References

About LibraryThing. Retrieved 10/18, 2009, from www.librarything.com/about

Buckland, M. (1999). Vocabulary as a central concept in library and information science. T. Arpanac et al., (Eds.). *Digital libraries: Interdisciplinary concepts, challenges, and opportunities: Proceedings of the Third International Conference on Conceptions of Library and Information Science,* Dubrovnik, Croatia, 23-26 May 1999, 3-12. Zagreb: Lokve. Retrieved 11/06,2009 from http://www.sims.berkeley.edu/~buckland/colisvoc.htm

Cattuto, C. (2008). Emergent Community Structure In Social Tagging Systems. *Advances in Complex Systems, 11* (4), 597-608.

Dye, J. (2006). Folksonomy. *EContent, 29* (3), 38-43.

Fry, E. (2007). Of torquetums, flute cases, and puff sleeves. *Art Documentation V.26 no.1 (Spring 2007) P.21-7, 26* (1), 21$l27-21$l27.

Furner, J. (2008). User tagging of library resources. *International Cataloguing and Bibliographic Control V.37 no.3 (July/September 2008) P.47-51, 37* (3), 47$l51-47$l51.

Golder, S.A. and Huberman, B.A. (2005). *The structure of collaborative tagging systems.* Retrieved 10/17, 2009, from http://www.hpl.hp.com/research/idl/papers/tags/tags.pdf

Kroski, E. (2007). Folksonomies and user-based tagging. In N. Courtney (Ed.), *Library 2.0 and beyond* (pp. 91-103). Westport, CT: Libraries Unlimited.

Macgregor, G. (2006). Collaborative tagging as a knowledge organisation and resource discovery tool. *Library Review, 55* (5)

McElfresh, L. K. (2008). Folksonomies and the future of subject cataloging. *Technicalities V.28 no.2 (March/April 2008) P.3-6, 28* (2), 3$l6-3$l6.

Mendes, L. H. (2009). Subjecting the catalog to tagging. *Library Hi Tech, 27* (1)

Pera, M. S. (2009). Sophisticated library search strategy using folksonomies and similarity matching. *Journal of the American Society for Information Science and Technology V.60 no.7 (July 2009) P.1392-406, 60* (7), 1392$l1406-1392$l1406.

Peterson, E. (2006). Beneath the metadata. *D-Lib Magazine V.12 no.11 (November 2006) P.1, 12* (11), 1.

Peterson, E. (2008). Parallel systems. *Library Philosophy and Practice V.2008 (2008) P.1-5, 2008,* 1$l5-1$l5.

Peterson, E. (2009). Patron preferences for folksonomy tags. *Evidence Based Library and Information Practice V.4 no.1 (2009) P.53-6, 4* (1), 53$l56-53$l56.

Rolla, P. J. (2009). User tags versus subject headings. *Library Resources & Technical Services V.53 no.3 (July 2009) P.174-84, 53* (3), 174$l184-174$l184.

Sanders, D. (2008). Tag--you're it! *American Libraries, 39* (11), 52-54.

Shirky, C. *Ontology is overrated: Categories, links, and tags.* Retrieved 10/17, 2009, from http://shirky.com/writings/ontology_overrated.html

Smith, G. (2008). *Tagging: People-powered metadata for the social web.* Berkeley, Calif: New Riders. Retrieved from http://proquest.safaribooksonline.com.libaccess.sjlibrary.org/?uiCode=csusanjose&xmlId=9780321550149

Snipes, P. R. (2007). Folksonomy vs. minnie earl and melville. *Library Media Connection V.25 no.7 (April/May 2007) P.54-6, 25* (7), 54$l56-54$l56.

Snuderl, K. (2008). Tagging. *Statistical Journal of the IAOS, 25* (3), 125-132.

Spiteri, L. (2006). The use of folksonomies in public library catalogues. *Serials Librarian,* 51(2), 75-89.

Spiteri, L. F. (2007). Structure and form of folksonomy tags. *Information Technology and Libraries V.26 no.3 (September 2007) P.13-25, 26* (3), 13$l25-13$l25.

Steele, T. (2009). New cooperative cataloging. *Library Hi Tech V.27 no.1 (2009) P.68-77, 27* (1), 68$l77-68$l77.

Suster, M. (2006). Folksonomy. *AIIM E-Doc Magazine V.20 no.6 (November/December 2006) P.20-1, 20* (6), 20$l21-20$l21.

Thomas, M. (2009). To tag or not to tag? *Library Hi Tech, 27* (3)

Trant, J. (2009). Tagging, folksonomy and art museums. *JODI: Journal of Digital Information, 10* (1), 5-5.

Vander Wal, T. (2005). *Folksonomy definition and Wikipedia.* Retrieved October 17, 2009 from http://vanderwal.net/random/entrysel.php?blog=1750

West, J. (2007). Subject headings 2.0. *Library Media Connection V.25 no.7 (April/May 2007) P.58-9, 25* (7), 58$l59-58$l59.

Digital Postcard Collections: Consistency and Retrieval

Lynne Noone

Lynne Noone can be reached at: noones1@pacbell.net

Introduction

There are many institutions with large digital collections of local historical significance for particular areas of location, culture and events. Different metadata schemes are used to describe these digital objects throughout the various institutions that maintain these collections. Dublin Core is a commonly used scheme to describe many of these collections. While some institutions choose to create their metadata locally, many universities and digital archives use CONTENTdm to help in the creation and display of metadata records to describe the works in their collections. Many of these institutions then share their records through a large repository called OAIster. While Dublin Core offers flexibility in the creation of metadata used to describe digital objects, how effectively this metadata achieves interoperability is an important issue in subject searching. The question of how interoperable subject and location search terms are in five different digital libraries using CONTENTdm is discussed in this paper.

Statement of the Problem

Digital library collections are created for the use of both the general public and academic institutions alike. Having the ability to create metadata records locally, allows the

creator of the collection the flexibility to design schemes that reflect the local nature of specific objects in their collection. Working in CONTENTdm, allows libraries to create records for objects on a local level using CONTENTdm templates and customizing them to meet their local needs. CONTENTdm libraries can share their metadata through OAIster, a repository of metadata records, now searchable through WorldCat. When searching outside the library's local collection for these items, the choice of vocabulary of subject headings and the fields chosen to display location and format of these local collections may inhibit interoperability despite the fact that the intent of these libraries is just the opposite.

Background

CONTENTdm

CONTENTdm is software that handles storage, management and delivery of digital collections. It is managed by OCLC (OCLC CONTENTdm overview, website). CONTENTdm is a way for libraries to quickly and easily create metadata and store digital collections either on their own servers with CONTENTdm software or on the server of OCLC. Many libraries, like the University of South Carolina, choose CONTENTdm for its ease of use and stable support. It also is an attractive solution to the building of a digital collection for those institutions that do not have the resources to create digital collections without this type of support (Swain, 2006, 58).

CONTENTdm allows for a variety of controlled vocabularies, as the software offers 10 integrated thesauri to choose from. The creator of metadata for a collection has the option to use his or her own designed vocabulary as well (OCLC CONTENTdm collection building and management, website). CONTENTdm is flexible, so that a variety of data standards such as XML, Dublin Core and METS that can be used in building a collection (OCLC CONTENTdm collection building and management, website).

Open Archives Initiative

The Open Archives Initiative (OAI) is an initiative that aims to "develop and promote interoperability standards that aim to facilitate the efficient dissemination of content" (Open Archives Initiative, website). With the OAI, different types of metadata can be harvested from different repositories making access to information easier for the user. The OAI, aims to promote interoperability standards that will aid in the effective retrieval of digital information (Open Archives Initiative, website).

The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) came out of the OAI. The OAI-PMH helps to bring data together in one place to enable the searching of the data at the same time. It defines a mechanism for harvesting metadata from different repositories (Open Archives Initiative, website).

OAIster

OAIster is a union catalog that contains millions of records harvested from open archive collections using the OAI-PMH. OAIster is currently managed by OCLC and is available for searching through WorldCat (OCLC. The OAIster database, website).

Users of CONTENTdm have the option to upload their records to WorldCat to lend themselves to discovery by users searching through WorldCat. (OCLC. CONTENTdm, overview. Website).

Literature Review

Interoperability

While digital library projects and organizations are plentiful, so too are the problems that exist surrounding them. With all the metadata that is being created, there is no one standard that everyone who creates metadata uses. While XML is a clear choice to use in creating metadata records to allow for flexibility in how the data is described, there is still the problem of interoperability. Interoperability, as described by the NISO, is "the ability of multiple systems with different hardware and software platforms data structures and interfaces to exchange data with minimal loss of content and functionality". (NISO, 2004, 2). Interoperability is amongst the most important issues in creating good metadata records (Chan, Zeng, 2006, 3).

One of the great things about metadata is the way that it can be manipulated depending on the type of collection it is describing. At the same time, the flexible quality of metadata can be a negative for the user when trying to do an integrated search in a repository of records created with different schemes. A user may assume that all relevant records from different collections will be retrieved with a particular search term, however, if two collections describe similar elements differently from each other (through choice of different fields for locations or vocabulary for subject headings for example), relevant records may be missed.

Throughout the process of the digitization of library cataloging records, MARC (Machine-Readable Cataloging) has been the standard metadata used. This standard format for cataloging records has been a consistent and successful method for library systems (Chan & Zeng, 2006, 4). For digital libraries, however, this approach to standards and uniformity does not work in describing different digital objects. Different schemes are created to service different needs and audiences, and a one-size fits all standard does not work (NISO, 2004, 13).

One way for digital libraries to achieve maximum interoperability would be for all collections to adopt the same controlled vocabulary for subject headings rather than amend the vocabulary to reflect their local collection (Nicholson, Shiri, 2003, 58). As stated above, however, objects from different digital collections have different characteristics that, in order to be described accurately must have the flexibility of describing them with vocabulary that is appropriate to the item. With so many different types of objects being described, a variety of different vocabularies have been developed to describe them: Library of Congress Subject Headings (LCSH) and the Art and Architecture Thesaurus and Medical Subject Headings to name a couple. This inconsistency in subject heading creation leads to the issue of how best to approach the issue of interoperability for subject access (Mitchell, 2006, 20).

Another consideration in the discussion of interoperability is the role of the user. While the library community has a certain comfort level with LCSH, the average user will search with a more general term or keyword (Mitchell, 2006, 21). If metadata records are being designed for ease of user searching, then the rigidity of LCSH is counterintuitive to the concept of a user-centered system. Lois Mai Chan and Theodora Hodges address this issue in 2000 and suggest that a simpler form of LCSH is necessary for the future. (Chan,

Hodges, 225, 2000). There already exist a variety of vocabularies to describe different types of materials. For example, the Thesaurus of Graphic Materials (TGM) was created as "a tool for indexing visual materials by subject and genre/format" (Library of Congress. Thesaurus for Graphic Materials II. Website). Chan and Hodges, however, are looking at newer and even simpler options. Faceted Application of Subject Terminology (FAST) is one type of vocabulary that has been created by "adapting the LCSH with a simplified syntax…to retain the very rich vocabulary of LSH while making the schema easier to understand, control, apply and use" (OCLC. FAST. Website).

In traditional cataloging, librarians are trained to create records using MARC, LCSH and Authority Control. By taking great care in record creation, library OPACs have achieved interoperability for books and other media that has been entered into library OPACs and these records are successfully shared through federated catalogs like WorldCat. Digital collections, however, have different considerations than traditional library collections do. For one thing, digital collections describe a variety of objects that require vocabulary specific to the object in order to describe it. Also, librarians do not necessarily create metadata records for digital collections trained in LCSH and authority control, but it is often students and paraprofessionals. Vocabularies like FAST may help to create some consistency in subject heading creation. It will be interesting to observe whether vocabularies like FAST will become a standard in the creation of metadata records for digital resources as FAST is designed to be used by people without extensive training as it is easy to use, understand and maintain (Dean, 2004, 333).

Research Questions

- How important is the role of controlled vocabulary in metadata descriptions of objects for facilitating successful searching within a collection? Within a repository (OAIster)?
- Does the field where location information and item type information are placed make a difference in a user's ability to access information?

Secondary Question

- How general or how specific should metadata descriptions of objects be to facilitate successful searching within a collection?

Methodology

Five digital libraries currently using CONTENTdm and sharing their collections through OAIster were identified to survey. CONTENTdm libraries displaying their collection in OAIster were selected, as their choice to display their records in OAIster implies the library's desire to share their records with a broad range of users. Also, many digital library collections choose CONTENTdm to help create their digital collections and having all the libraries using the same database software allows for some level of uniformity. From each library, three postcards in the collection were randomly selected for analysis. A comparative analysis was done using the following criteria:

- In the main search field of the collection, perform a search using the term "postcards" to determine the results produced using this term. After the search is conducted, note whether or not there is a way to narrow numbers within the results.

- What field is location defined in? What controlled vocabulary (if any) is used for subject headings? (e.g. Library of Congress Subject Headings, some form of local subject headings).
- Is "postcard" or "postcards" defined in the field "Type"?
- Is "postcard" or "postcards" defined in the field "Format"?
- Is postcard listed as a subject?
- When running a Boolean search with "postcard" and "location" in the digital collection, how many results are retrieved?
- What are the results of the same Boolean search in OAIster?

Results

Diagram R1

Santa Clara University	Image 1	Image 2	Image 3
Postcard search # retrieved	117	117	117
Refineable Y/N	N	Y	Y
Location defined in	title	title	title
CV subject headings Y/N	Y	Y	Y
Controlled Vocabulary Type	TGM*	TGM*	TGM*
"Postcard" defined in "format" Y/N	N	N	N
"Postcard" defined in "type" Y/N	Y	Y	Y
Postcard listed as subject Y/N	Y	Y	Y
Boolean search in DC # retrieved	54	90	117
Postcard retrieved in DC Y/N/U	Y	Y	Y
Boolean search in OAIster # retrieved	38	202	550
Postcard retrieved in OAIster Y/N/U	N	U	U
Y=yes N=no U=undetermined			
CV = Controlled Vocabulary			
DC = Digital Collection			
Boolean search terms are the Location as indicated on the postcard record plus the term "postcard".			
Search is run using the advanced search function. If no advanced search function is available, run the search using the Boolean term AND between the two terms in the available search box			
TGM = Thesaurus for Graphic Materials			
Notes: *TGM with local geographic additions			

Diagram R2

University of Miami	Image 1	Image 2	Image 3
Postcard search # retrieved	369	369	369
Refineable Y/N	Y	Y	Y
by:	date	date	date
	creator	creator	creator
Location defined in	state	state	state
	county	county	county
	city	city	city
CV subject headings Y/N	N	N	N
Controlled Vocabulary Type	local	local	local
"Postcard" defined in "format" Y/N	N	N	N
"Postcard" defined in "type" Y/N	N	N	N
Postcard listed as subject Y/N	Y	Y	Y
Boolean search in DC # retrieved	42	272	273
Postcard retrieved in DC Y/N/U	Y	Y	U
Boolean search in OAIster # retrieved	47	274	388
Postcard retrieved in OAIster Y/N/U	Y	U	U
Y=yes N=no U=undetermined			
CV = Controlled Vocabulary			
DC = Digital Collection			
Boolean search terms are the Location as indicated on the postcard record plus the term "postcard".			
Search is run using the advanced search function. If no advanced search function is available, run the search using the Boolean term AND between the two terms in the available search box			
TGM = Thesaurus for Graphic Materials			
Notes:			

Diagram R3

University of Louisville	Image 1	Image 2	Image 3
Postcard search # retrieved	630	630	630
Refineable Y/N	N	N	N
Location defined in	location depicted	location depicted	location depicted
CV subject headings Y/N	Y	Y	Y
Controlled Vocabulary Type	TGM	TGM	TGM
"Postcard" defined in "format" Y/N	N	N	N
"Postcard" defined in "type" Y/N	Y	Y	Y
Postcard listed as subject Y/N	N	N	N
Boolean search in DC # retrieved*	8	1	26
Postcard retrieved in DC Y/N/U	Y	Y	y
Boolean search in OAIster # retrieved	70	50	36
Postcard retrieved in OAIster Y/N/U	Y**	Y	Y
Y=yes N=no U=undetermined			
CV = Controlled Vocabulary			
DC = Digital Collection			
Boolean search terms are the Location as indicated on the postcard record plus the term "postcard".			
Search is run using the advanced search function. If no advanced search function is available, run the search using the Boolean term AND between the two terms in the available search box			
TGM = Thesaurus for Graphic Materials			
Notes: It was necessary to perform a search using exact fields (most users would not know how to do this) **Also retrieved from Auburn University who had "postcards" in subject heading. Auburn's listing came earlier in the results			

Diagram R4

University of Washington	Image 1	Image 2	Image 3
Postcard search # retrieved	522	522	522
Refineable Y/N	N	N	N
Location defined in	publisher location	location depicted	places
CV subject headings Y/N	Y	Y	Y
Controlled Vocabulary Type	TGM	TGM	TGM
"Postcard" defined in "format" Y/N	N	N	N
"Postcard" defined in "type" Y/N	N	N	N
Postcard listed as subject Y/N	Y	Y	Y
Boolean search in DC # retrieved	543	521	521
Postcard retrieved in DC Y/N/U	Y	Y	U
Boolean search in OAIster # retrieved	267	759	759
Postcard retrieved in OAIster Y/N/U	U	U	U
Y=yes N=no U=undetermined			
CV = Controlled Vocabulary			
DC = Digital Collection			
Boolean search terms are the Location as indicated on the postcard record plus the term "postcard".			
Search is run using the advanced search function. If no advanced search function is available, run the search using the Boolean term AND between the two terms in the available search box			
TGM = Thesaurus for Graphic Materials			
Notes:			

Diagram R5

University of South Carolina	Image 1	Image 2	Image 3
Postcard search # retrieved	80	80	80
Refineable Y/N	N	N	N
Location defined in	subject	subject	subject
CV subject headings Y/N	N	N	N
Controlled Vocabulary Type	local	local	local
"Postcard" defined in "format" Y/N	N	N	N
"Postcard" defined in "type" Y/N	N	N	N
Postcard listed as subject Y/N	Y	Y	Y
Boolean search in DC # retrieved	2	1	64
Postcard retrieved in DC Y/N/U	Y	Y	Y
Boolean search in OAIster # retrieved	6	2	380
Postcard retrieved in OAIster Y/N/U	N	N	U
Y=yes N=no U=undetermined			
CV = Controlled Vocabulary			
DC = Digital Collection			
Boolean search terms are theLocation as indicated on thepostcard record plus the term "postcard".			
Search is run using the advanced search function. If no advanced search function is available, run the search using the Boolean term AND between the two terms in the available search box			
TGM = Thesaurus for Graphic Materials			
Notes: in OAIster tried also South Carolina tried also Dillon County (S.C.)			

Discussion

Boolean Search in Digital Collection (Diagram D1)

Postcards in Collection	Boolean Results	Refined by	Percentage
U of Washington			
522	543	-21	-0.04%
522	521	1	na
522	521	1	na
U of South Carolina			
80	2	78	98%
80	1	79	99%
80	6	74	93%

U of Miami			
369	42	327	89%
369	272	97	26%
369	272	97	26%
U of Louisville			
630	8	628	99%
630	1	629	100%
630	26	604	96%
Santa Clara University			
117	54	63	54%
117	90	27	23%
117	117	0	0%

Comparison of Digital Collection Retrieval and OAIster Retrieval in Boolean Searches (Diagram D2)

	Digital Collection	OAIster	Difference	% More in OAIster
U of Washington				
Image 1	42	47	5	12%
Image 2	272	274	2	<1%
Image 3	273	388	115	42%
U of South Carolina				
Image 1	2	6	4	50%
Image 2	1	2	1	100%
Image 3	64	380	316	20%
U of Miami				
Image 1	42	47	5	12%
Image 2	272	274	2	<1%
Image 3	273	388	114	42%
U of Louisville				
Image 1	8	70	62	13%
Image 2	1	50	49	>100%
Image 3	26	36	10	38%
Santa Clara University				
Image 1	54	38	-16	-29%
Image 2	90	202	112	80%
Image 3	117	550	433	27%

Although the sample of records for this study is small, there are a few things that stand out and lend themselves to further examination. While running the Boolean searches, it appears that creation of metadata at the local level has a direct impact on users searching in a repository like OAIster (through WorldCat). Using vocabulary to describe an object with

terms specific to its local nature may become a barrier to retrieval for someone unfamiliar with these terms. Take for example the University of South Carolina's digital collection (Diagram R5). The terms used for subject headings do not fall under LCSH, TGM or any other recognizable vocabulary. One of the terms used in subject headings is "Latta SC--Pictorial work" (Appendix 5. University of South Carolina, Image 2), which, unless a user is familiar with what or where Latta SC is, is too obscure of a search term for the general public. However, there are other terms in the subject headings for this image that are useful for retrieval, so the creator of the subject terms was thoughtful in creating the heading. Keeping the more local subject heading does have its purpose in describing the image, but does not provide help in retrieval.

The other barrier to user discovery is the choice of *field* where the location information is placed. If we look again at the University of South Carolina, general location information (i.e. state) is in the subject area, and more specific location information is in two separate data fields of county and region (Appendix 5. University of South Carolina image 1, image 2, image 3). In running a Boolean search for the University of South Carolina, it is difficult to determine what to use for a location term. If a specific region is used, there are a high percentage of relevant results (Diagram D1). The same is true for searching in OIAster, however, if using location as a search term in OAIster, one would have to know the exact county or region information to find any records for these particular subjects. If the user searches with the location using a more general term like "South Carolina", the results would not have been so precise and most likely thousands of records would have been retrieved.

In looking at the University of Miami, (Diagram R2) there are also problems arising from the field where the location is defined. As with the University of South Carolina, the location is defined in three separate fields, in this case: state, county and city. It is also difficult to determine which definition of location to search with for the Boolean search. In the case of Image 1 (Appendix 2. University of Miami, Image 1), the Boolean search using terms from this image results in more precise results than using terms from images 2 and 3 (Appendix 2. University of Miami, Image 2 and 3). As with the University of South Carolina, it is necessary to be familiar with the regional locations of the postcards in order to achieve this high level of precision with searching locations. When the same search is run in OAIster, the number of retrieved records is similar. This may indicate that if location search terms are too specific, they do not allow for the retrieval additional relevant (but not to specific location) records.

In the comparison of Boolean searching using "postcard" and location information done in each digital collection and OAIster, the results vary widely within each sample group of postcards, and there is not enough data to reach any conclusions (Diagram D2). One of the reasons that the results vary so widely may be that the location terms are so specific to the collections, that it is hit or miss whether or not those terms would be used for description of works in another collection thereby resulting in greater retrieval. When location is identified using county and region as fields, these search terms can be missed as users tend to search using broader terms. When the location is more general, the percentage of results seems to increase, however, this analysis is not within the scope of this project.

Postcard Locations (Diagram D3)

	Number	Percentage
state	3	20
county	3	20
city	3	20
location depicted	4	27
title	3	20
publisher location	3	20
subject	3	20
places	1	7

How location is defined is important to the retrieval of records for a search (Diagram D3). If the location is defined too broadly, like the University of Washington's collection, for example United State--Washington—Seattle (Appendix 4. University of Washington, Image 2), it retrieves results from beyond Seattle. However, when the location is more specific, as in the University of South Carolina, which is region and county specific (Appendix 5. University of South Carolina, Image 1), unless you know exactly the name of these counties, you would find it difficult to retrieve any records from these collections if you searched with locations. In the University of Washington (Appendix 4. University of Washington, Image1, 2 and 3) it was difficult to even find a postcard in the sample group by doing a search with location. As noted in Diagram D3, there is no one particular field used to describe "location" in any of the libraries studied. Though the scope of this study was small, it is possible to infer that digital collections in general, have a variety of ways to display "location", and this disparity of location description does not aid with interoperability.

Postcard Fields (Diagram D4)

	Number	Percent
Postcard defined in "format"	0	0
Postcard defined in "type"	6	40%
Postcard defined in "subject"	12	80%
Postcard defined in more than one field	3	20%

In looking at how "postcard" is defined, in four of the five collections looked at, the creator of the metadata chose to define "postcard" in the subject field. Format was not chosen by any of the creators to define postcard, most likely because the metadata is describing the representation of the object in digital form, not the object itself. Only one library in the study chose to define postcard in the "type" field (Diagram D4). It is noteable that there is a consensus in how the work is described as a representation of the work. No matter how "postcard" is defined (by type, in subject etc), records are retrieved when doing a general search with "postcards". How the creator of the metadata chooses to define "postcard" does not seem to effect retrieval of postcard records.

Refineable After Initial Search (Diagram D5)

	Number	Percentage
Yes	3	20
No	12	80

Whether or not a set of retrieved records is refineable can be useful to a user. In looking at the collections, there was only one library that allowed for refinement after a search was completed, namely the University of Miami, leaving 80% of the libraries examined in this study with no way to refine a search (Diagram D5). This is frustrating on a certain level, since if a search does not retrieve desirable results, the user must go back to the search field and try again. If an option to refine a search presents itself to the user, the user is given ideas of how further to refine his search based on fields that the user might not have considered. This is a useful feature of a database to users, especially if users are a consideration in how the metadata is created and searched for. One need only look to many library OPACs and even the search interface on many online retail sites like Amazon.com and Barnes and Noble to see this as an effective feature to a database interface.

While postcard is not listed in the "format" field in any of the records examined (Diagram D4), format is one of the ways to refine a search in OAIster. However, since "format" of the object is not defined as "postcard" for any of the postcards looked at for this study, this option to refine a search in OAIster is irrelevant here. How postcards are defined in a metadata record most likely does not inhibit searching on OAIster, although it is hard to definitively conclude this (Diagram D2).

Controlled Vocabularies Used to Describe Subject Headings (Diagram D6)

Type	Number	Percentage
Local	6	40%
TGM	9	60%

In looking at the different vocabularies that are used, as expected, there are a variety of vocabularies. Most notably missing is LCSH. It is difficult to tell how these vocabularies affect interoperability, as localization of the subject headings make it difficult to make similar comparisons. If users are searching with keywords for subjects, however, and not controlled vocabulary, it is unclear whether or not the controlled vocabulary in the subject headings makes a difference. It seems that while controlled vocabulary is important for consistency of recall, the information that the local subject headings contain are important for discovery of information about the work being described. One thing that is clear in looking at the different vocabularies used for each collection is that digital collections have a tendency to rely on local subject headings to describe the collection. How subjects are defined in OAIster is where interoperability becomes an issue. It is difficult to define a measurement for this. More research would need to be done to make definitive conclusions about how choice of field for the term "postcard" effects search retrieval. How the different controlled vocabularies affect interoperability is not clear from the research performed for this paper.

Challenges

Looking at the variety of subject headings and fields that location was defined in proved to be challenging in trying to find some way to analyze how these fields aided or hindered access and interoperability. Without consistent fields used for location and the same vocabulary used for subject headings, comparisons were difficult. Because of the flexibility of metadata, the challenge lied in finding common ground to analyze. The original idea of this project was to compare the benefits of using one metadata scheme over another in describing postcards in digital collections, however, it is difficult to find actual metadata code to look at. Many digital libraries use CONTENTdm to create, manage and store their records, so there was no way to view the source code. Also, because of the flexible nature of metadata, it is difficult to compare the creation of one collection with another.

It was also difficult to determine the effectiveness of subject headings without having an idea of how a user might conduct a search and what knowledge a user possesses at the onset of a search.

Conclusion

While there is a consistent use of the subject fields in the postcard metadata records studied, the vocabulary within these fields is not consistent. Also, there is an inconsistent use of fields used to describe locations of the postcards within each collection. Finally, in the collections studied, there is a tendency to describe the postcards as digital object representations rather than as actual postcards, however, in describing "postcard", there are a variety of fields where this term is listed. This is in line with Cole and Shreeve's findings that after creators of metadata determine whether they will describe the object in the collection as a representation of the work, or as the work itself, the term used to describe the work is not held by any standard (Cole and Shreeve, 2004, 175).

Perhaps the key is not to focus on if local subject headings are effective for record retrieval, but rather, the focus should be on determining if there are other fields that should have more standardization such as format, type and location in order to help in retrieval. Focusing on how particular formats are defined, and in what fields location information is created can be useful in helping users find relevant records. The information in the local subject headings is helpful for the user once a record is found and can help in locating more information and learning more about an item.

Location fields are also important for interoperability and making location information more consistent and not too narrow in definition can help with interoperability for the end user. While the flexibility of metadata creation allows for a rich content of local information to be available to the user, it must be findable in order to be useful. Digital objects that are customized locally are not necessarily optimized for retrieval.

This project initially set out to find if it is more useful for subject headings and location information to have general or specific descriptions. There is a thin line between having too broad of a topic so that users retrieve too many records and too narrow of one where a user may never find the record at all. If a subject heading is so deconstructed as to become so generic users may not be able to find records limited to a place (Qiang, J. 2008, 108). If the subject heading is too specific, however, unless a user knows exactly the local term to search, they many never find information. Although this paper does not set out to analyze user interfaces, refineable searches can be useful in helping a user better define their topic. Users are accustomed to interfaces from library catalogs and retail databases

that offer suggestions for searching records with similar search terms they have used. Finally, the concept of creating controlled vocabularies based on already existing vocabularies (like FAST) is also worth exploring further.

As more and more digital collections are available for the general public, the impact of subject heading vocabulary design and creating consistent fields for location information and description of a format for the work being represented will continue to be explored as more people see the potential to accessing the digital collections that exist and metadata creators find more ways to successfully achieve interoperability.

References

Chan, L., & Hodges, T. (2000). *Entering the millennium:* a new century for LCSH. *Cataloging & Classification Quarterly, 29* (1/2), 225-34. Retrieved November 25, 2009, from Library Lit & Inf Full Text database.

Chan, L., & Zeng, M. (2006). Metadata Interoperability and Standardization - A Study of Methodology, Part I: Achieving Interoperability at the Schema Level. *D-Lib Magazine, 12* (6), p. 1. Retrieved November 16, 2009, from Library Lit & Inf Full Text database.

Cole, T.W. & Shreeves, S. L. *Lessons learned from the Illinois OAI Metadata Harvesting Project.* Retrieved on Wednesday, November 25, 2009 from http://books.google.com/books?id=2nzLQHn1WAUC&printsec=frontcover&source=gbs_v2_summary_r&cad=0#v=onepage&q=&f=false

Dean, R. (2004). FAST: Development of Simplified Headings for Metadata. *Cataloging & Classification Quarterly, 39* (1/2), 331-52. Retrieved November 29, 2009, from Library Lit & Inf Full Text database.

Dragon, P. (2009). Name Authority Control in Local Digitization Projects and the Eastern North Carolina Postcard Collection. *Library Resources & Technical Services, 53* (3), 185-96. Retrieved November 15, 2009, from Library Lit & Inf Full Text database.

Hillmann, D. (2008). Present at the Creation. *Technicalities, 28* (3), 7-9. Retrieved November 15, 2009, from Library Lit & Inf Full Text database.

Library of Congress. Thesaurus for Graphic Materials II. Website. Retrieved Sunday on Sunday, November 22, 2009 from http://www.loc.gov/rr/print/tgm2/

Mitchell, N. (2006). Metadata basics: A literature survey and subject analysis. *The Southeastern Librarian, 54* (3), 18-24. Retrieved November 24, 2009, from Library Lit & Full Text database.

Nicholson, D., & Shiri, A. (2003). Interoperability in subject searching and browsing. *OCLC Systems & Services, 19* (2), 58-61. Retrieved November 24, 2009, from Library Lit & Inf Full Text database.

NISO (2004). Understanding metadata. *National Information Standards Organization.* NISO Press. Bethesda, MD. Retrieved November 16, 2009, from www.niso.org/standards/resources/UnderstandingMetadata.pdf

NISO Metadata Principle 1. Webpage. Retrieved Thursday November 19, 2009 from http://framework.niso.org/node/38

OCLC. CONTENTdm Overview. Website. Retrieved Sunday, November 21, 2009 from http://www.oclc.org/contentdm/overview/default.htm

OCLC. FAST. Website. Retrieved Sunday, November 21, 2009 from http://www.oclc.org/research/activities/fast/default.htm

OCLC. The OAIster Database. Website. Retrieved Sunday November 22, 2009 from http://www.oclc.org/oaister/

Open Archives Initiative. Website. Retrieved Sunday November 20, 2009 from http://www.openarchives.org/documents/FAQ.html#What%20is%20the%20mission%20of%20the%20Open%20Archives%20Initiative

Open Archives Initiative Metadata Harvesting Project. Website. Retrieved Monday November 16, 2009 from http://uilib-oai.sourceforge.net/

Qiang, J. (2008). Is FAST the Right Direction for a New System of Subject Cataloging and Metadata?. *Cataloging & Classification Quarterly, 45* (3), 91-110. Retrieved November 27, 2009, from Library Lit & Inf Full Text database.

Swain, S. (2006). University of South Carolina, CONTENTdm, and the Ege Leaves. *The Southeastern Librarian, 54* (1), 58-9. Retrieved November 22, 2009, from Library Lit & Inf Full Text database.

Library Postcard Collections Used for Project

Appendix 1

Santa Clara University Digital Collections, Projects and Initiatives

http://cms.scu.edu/library/collections/digital/

Image 1:

http://content.scu.edu/cdm4/item_viewer.php?CISOROOT=/svhocdm&CISOPTR=449&CISOBOX=1&REC=1

Image 2:

http://content.scu.edu/cdm4/item_viewer.php?CISOROOT=/svhocdm&CISOPTR=472&CISOBOX=1&REC=17

Image 3:

http://content.scu.edu/cdm4/item_viewer.php?CISOROOT=/svhocdm&CISOPTR=1468&CISOBOX=1&REC=16

Appendix 2

University of Miami Libraries Digital Initiatives

http://merrick.library.miami.edu/

Image 1:
http://merrick.library.miami.edu/cdm4/item_viewer.php?CISOROOT=/asm0299&CISOPTR=676&CISOBOX=1&REC=1

Image 2:

http://merrick.library.miami.edu/cdm4/item_viewer.php?CISOROOT=/asm0299&CISOPTR=596&CISOBOX=1&REC=21

Image 3:

http://merrick.library.miami.edu/cdm4/item_viewer.php?CISOROOT=/asm0299&CISOPTR=595&CISOBOX=1&REC=20

Appendix 3

University of Louisville Libraries Digital Collections

http://digital.library.louisville.edu/

Image 1:

http://digital.library.louisville.edu/cdm4/item_viewer.php?CISOROOT=/ulua001&CISOPTR=637&CISOBOX=1&REC=1

Image 2:

http://digital.library.louisville.edu/cdm4/item_viewer.php?CISOROOT=/ulua001&CISOPTR=263&CISOBOX=1&REC=15

Image 3:

http://digital.library.louisville.edu/cdm4/item_viewer.php?CISOROOT=/ulua001&CISOPTR=691&CISOBOX=1&REC=1

Appendix 4

University Libraries University of Washington Digital Collections

http://content.lib.washington.edu/

Image 1:

http://content.lib.washington.edu/cdm4/item_viewer.php?CISOROOT=/advert&CISOPTR=492&CISOBOX=1&REC=3

Image 2:

http://content.lib.washington.edu/cdm4/item_viewer.php?CISOROOT=/seattle&CISOPTR=1971&CISOBOX=1&REC=2

Image 3:

http://content.lib.washington.edu/cdm4/item_viewer.php?CISOROOT=/imlsmohai&CISOPTR=3039&CISOBOX=1&REC=6

Appendix 5

University of South Carolina Digital Collections

http://sc.edu/library/digital/index.php

Image 1:

http://digital.tcl.sc.edu/cdm4/item_viewer.php?CISOROOT=/rrc&CISOPTR=1689&REC=9

Image 2:

http://digital.tcl.sc.edu/cdm4/item_viewer.php?CISOROOT=/rrc&CISOPTR=1692&REC=5

Image 3:

http://digital.tcl.sc.edu/cdm4/item_viewer.php?CISOROOT=/bgp&CISOPTR=182&REC=7

A Barcode Scanning System for an Elementary School Library

Susan K. Genenbacher

Susan K. Genenbacher is District Librarian of for Southeastern 337 School District in Illinois. She has taught Language Arts to 6th, 7th, and 8th graders for 16 years, and is now working on Media Specialist certification. This project was necessary to bring students into the 21st century. It provides a tool that connects the students to the library and allows them to connect to other libraries through interlibrary loan. She can be reached at: sgenenbacher@southeastern337.com

Introduction

Southeastern School District was formed in 1970. The library presence was and still is very important to students, faculty, and school board. The library director has more than 30 years of knowledge, which shows quickly when students come in to search for a

particular book, or just want to read. Her conversation shows her interest in their needs, and she can quickly put several books in their hands.

In March of 2006, Southeastern High School experienced a fire that destroyed the school. The library suffered smoke and water damage. A company was hired to come in and pack up and dry and clean our books. They were able to save about 75 percent of the collections. The school year was finished in several local churches. Trailers were set up for the following year. The library was set up in two classrooms. In was very crowded and inconvenient but we survived. New books were ordered and during the summer of 2008 the library was ready to move into.

The new library is beautiful. When school started, the library still had manual check-out, with everything done by hand. With such a state of the art school with technology in every classroom and most of our books represented in the Alliance Library System, how could we not be automated? We used some of the funds received to rebuild the school after the fire. It took 6 weeks to get the rest of the books into the system. The elementary school was consolidated from two buildings to one. Library books were merged into one collection, which presented a good opportunity to weed.

The library also needed an automated circulation system, using barcodes and scanners, which would save time and money and be more accurate. Such a project would take about three month to complete. It was proposed that two computers be purchased so students can search electronically in the library. The project was started in the fall of 2008 and put to use in the second semester. The operation went rather smoothly and the Alliance came in for training after several weeks into the program.

Project Method

1. Contact the Alliance Library System to discover what must be done to complete the project.

2. Get prices on equipment and supplies, such as computer, scanner, barcodes, and student Ids.

3. Order barcodes for books and ID labels for students.

4. Contact technical support to download program.

5. Place barcodes on books and shelf cards.

6. Set up parameters for circulation program (checkout periods, fines, and so on.)

7. Train staff to use circulation program.

8. Begin barcoding books,

9. Notify Alliance Library System when all books are barcoded.

10. Enter student ID numbers into computer to create patron records.

11. Have system turned on.

Project Goal

The project goal was to have the scanning system up and running by second semester. Students will spend less time searching for information and be able to check out books more quickly, and an accurate record of overdues can be kept. Elementary students are demanding users and require a lot of the librarian's time. A computerized circulation system and computer access will give the librarian more time to spend with students.

Project Background

The district high school library collection had already been barcoded, and computers and scanners installed. Most of the elementary school books were already in the Alliance Library System database. The new elementary school was a state of the art building, with two computer labs with 60 computers and 50 laptops for student use, and a smart board in every classroom including the library but the library did not have automated circulation. How could this have been overlooked? To solve this problem, we purchased the SirsiDynix system from the Alliance system. We could add circulation information about the elementary school to the Alliance database for $500.00. After a computer and software were acquired for the project, the paper shelflist was used to identify books and barcode them. This took about six weeks, which meant that by December 15th we were using the new system. It was wonderful. My goal was to have the system up and running by second semester and it was. My second goal was to have the elementary library online by Christmas and it was going to happen.

Project Timeline

September 4: barcodes and scanner were ordered. I talked with the Alliance to set up our program. The technical person is coming in on September 18th to put program on the computer. 1 hour

September 9: Talked with Alliance to set the parameters for our program. 2 hour

September 21: barcodes arrive and I started putting them in the books and shelf cards. This week I have put in 40 hours and have about one third left to do.

September 28: The Book Fair is this week, but I was able to finish the barcodes. I put in 12 more hours on this.

October 5: This week I am putting the books in the system. I put in 55 hours and have about one third complete. I didn't realize how many easy books there are. I had a few problems with the Internet connection.

October 12: This week I am still putting in barcodes. I have experienced difficulty with the server going down. I have taken a drawer of shelf cards with me to the high school to work on them. I have to put everything in "stacks1" location so the Alliance can pull them out of my library at the high school. I have put in 35 hours this week.

October 19 th : Will continue putting the books into the system. It will take about four weeks to complete the barcoding. I put in 32 hours this week.

October 26: All the Easy Books and Biographies are in the system. The Library Aid helped put the books into the system. Together we have clocked 70 hours.

November 2: Started on the fiction today. We have parent conferences this week with no activity in the library. My aid and I clocked 60 hours and finished the fiction

November 9: We are starting the non-fiction this week My aid and I put in 40 hours this Week.

November 16: Finish the non-fiction this week. My aid and I put in 40 hours this week.

November 23: Fix problem books. Contact the Alliance to let them know I am finished with the barcodes. I put in 12 hours this week.

November 30th: Put Student IDs in the computer. It should take about 6 hours to complete.

December 1st: Call Alliance to let them know we are ready to scan. They will have to finish their program. It should take a couple of days.

December 7th: Start scanning books. The project started September 4, 2009 and will be completed by December 7, 2009. I have clocked about 330 hours and my aid has clocked about 105 hours on this project.

Consultation

I consulted with the librarian at Warsaw Public Library. She automated her library two years ago, and stated that it was very stressful but a worthwhile project. The city was building a new library and she wanted to be automated before it opened. She had asked the previous director to come back part-time to help run the library so she could concentrate on getting the collection into the database. They have a very active Friends of the Library group. Two Friends members helped with barcoding and data entry. It took the three of them three months to complete this task. They were closed four days to complete the move to the new library, and when they opened, they were automated.

Problems Encountered

The most frustrating problem was an unreliable wireless Internet connection. I had to take a drawer to another part of the building and work when that happened. My aid and I also took several drawers to the high school and worked there while the connection problem was being worked on. There were also a few problems getting the program set up. I received valuable technical support from the Alliance system staff.

Conclusion

Being persistent in proposing this project and demonstrating the need for it paid off. The staff at the Alliance Library System were extremely helpful. They were patient and very generous with their time. Selling your library is up to you. If you want something, research the benefits, requirements, and cost. If you are prepared and confident, you may just get what you are asking for. In doing this project, I also became acquainted with the books in the collection. It gave me a better insight about weeding and purchasing new books.

PNLA Quarterly
322B Love Library
PO Box 884100
13th & R St
University of Nebraska—Lincoln
Lincoln NE 68588-4100