Metadata Effectiveness in Internet Discovery: An Analysis of Digital Collection Metadata Elements and Internet Search Engine Keywords

Le Yang
Assistant Librarian
Texas Tech University
le.yang@ttu.edu

Accepted: March 19, 2015

Anticipated Publication Date: January 1, 2016

Manuscript#: crl15-722

ABSTRACT

This study analyzed digital item metadata and keywords from Internet search engines, in order to find out what metadata elements actually facilitate discovery of digital collections through Internet keyword searching and how significantly each metadata element affects the discovery of items in a digital repository. The study found that keywords from Internet search engines matched values in eight metadata elements and resulted in landing visits to the digital repository. Findings of the study indicate that three specific metadata elements are effective in enhancing discoverability of digital collections through Internet search engines, including Dublin Core metadata elements Title, Description, and Subject.

INTRODUCTION

In recent decades, the rapid increase in the number of digital repositories has called for in-depth research on metadata quality evaluations. As Ann Windnagel claimed, the success of a digital repository depends largely on the quality of its metadata.[1] Discussions about metadata evaluation and evaluation criteria have lasted for more than a decade. Relevant issues including metadata definition, primary purpose, and functionality perspectives have all been examined. Although several evaluation standards of metadata quality exist, there is no one answer to the question of how to evaluate metadata quality.[2]

In the literature, discoverability issues are the focus receiving the most impact in several metadata evaluation standards. As items are put into digital repositories instead of the catalog, they move out of the controlled environment of the library. The metadata records, instead of just having to work within a system search, also have to be effective for external indexing by search engines in order to be found. Challenges of deploying metadata to the Internet to facilitate information discovery and retrieval have persisted for nearly twenty years; however, few studies have been conducted to test the effectiveness of metadata on online resource discovery by search engines. No relevant studies on metadata effectiveness between digital collections and search engines are found in the current literature.

This article aims to examine the implemented metadata of a digital collection and to evaluate the metadata effectiveness on digital resource discovery by Internet search engines. Utilizing Google Analytics on [insert institute's name] Libraries' digital collections, this study extracts the digital content that is frequently visited by search engine traffic, analyzes the valid keywords that are entered in search engines, and compares keywords against implemented values of metadata elements. Results of the study show that keywords from Internet search engines find matched values in certain metadata elements in the digital collections. The findings indicate that certain metadata elements are more effective in facilitating discovery of digital collections on Internet search engines and those creating metadata records should pay extra attention to these elements.

LITERATURE REVIEW

Since the 1990s, a variety of digital libraries have been established to store and provide access to digital resources. Deploying cataloging rules for digital resources has led to favoring metadata as the best means of describing and discovering resources on the Web.[3]

Mehdi Safari claimed that as a result of the exponential growth of information resources on the Internet, metadata expanded beyond the traditional environment to address comprehensive

issues involving effective resource description and discovery.[4] Jung-Ran Park emphasized that the rapid proliferation of digital repositories had called for research on metadata quality evaluation.[5] F. O. Ininkaye, A. B. C. Robert, and B. A. Ojokoh echoed that the rapid increase of digital repositories motivated research on metadata quality and measurement.[6] Windnagel stressed the importance of metadata quality in the usability of a digital repository, and asserted that the success of a digital repository depends in large part on the quality of its metadata.[7]

Metadata is a heavily used term for which different definitions have been offered. In general, it may be defined as structured data about data, which "characterizes source data, describes their relationships, and supports the discovery and effective use of source data."[8] Alternatively, metadata can be defined as a structured set of elements that describe the information resources for the purpose of identification, discovery, and use of information.[9] According to National Information Standards Organization (NISO), metadata is structured information that describes, explains, locates, or otherwise makes easier to retrieve, use, or manage an information resources.[10]

The conversation about metadata quality has lasted for more than a decade and has focused primarily on determining how general quality criteria might be established in an environment where diversity of metadata formats coexist.[11] In 1997, William E. Moen, Erin L. Stewart, and Charles R. McClure summarized 23 evaluation criteria from literature on metadata quality, and provided an in-depth discussion through examination of metadata records of 42 federal agencies. This comprehensive study specified four major sets of criteria including completeness, accuracy, consistency, and currency.[12] In 2004, Thomas R. Bruce and Diane I. Hillmann stressed that completeness of metadata could be measured by connections of individual objects to the parent collections, which reflect the functional purpose of metadata in resource discovery and use.[13] In 2009, Park examined the state of the art of metadata quality evaluation and summarized that the quality of metadata reflects the degree to which the metadata performs the core bibliographic functions including discovery, use, provenance, currency, authenticity, and administration. In other words, Park concluded that the principle purpose of metadata is to a large degree related to that of traditional online library catalogs and databases in finding, identifying, selecting, and obtaining items.[14]

Discoverability has been the key issue in the standard of metadata quality evaluation. As Sevim McCutcheon noted, ambitious digitization projects have made digitized books and articles increasingly discoverable and accessible to Web users via keyword searching.[15] Tyler E. Phelps pointed out that the challenge of bringing metadata-based order to the Internet to facilitate information retrieval has persisted for nearly twenty years.[16]

The major purpose of any metadata, said Warwick Cathro, is to facilitate and improve the retrieval of information for information analysis and predictive decision-making process.[17] Stuart Weibel et al. specified that resources discovery is the most pressing need that metadata can satisfy, and thus believed that a simpler metadata scheme, such as Dublin Core, is desired because such a schema presents the minimum number of metadata elements required to facilitate resource discovery on the Internet.[18] In the report, NISO stressed that an important reason for creating

descriptive metadata is to facilitate discovery of relevant information.[19] Descriptive metadata provides a way to facilitate easy searching, retrieval, and management of information resources.[20]

Safari expressed the concerns about metadata effectiveness, asking if metadata provides a basis for increased effectiveness of retrieval by search engines.[21] While there have been studies done to evaluate search engines, as Safari noted, few studies have been done to test the effectiveness of metadata on resource discovery by search engines. Moreover, these studies focus primarily on Web page resources instead of digital collections, and research conducted by different scholars in different times yield inconsistent results.[22]

One example on how embedded metadata effects retrieval of Web pages was conducted by Thomas P. Turner and Lise Brackbill, and found that the use of keyword meta tags made a significant improvement on the retrievability of a Web page.[23] Jin Zhang and Alexandra Dimitroff also claimed that since metadata techniques have been applied to various formats of digital resources, metadata should make Internet resources more organized, informative, searchable, and accessible, and the precision of information retrieved by search engines should improve substantially.[24] In their research, Zhang and Dimitroff found that Web sites with Dublin Core metadata were returned by search engines significantly higher and faster in the search list than those sites without. Such findings demonstrate that implementation of metadata effectively improves the visibility of search results lists in search engines. [25]

In 2001, Robin Henshaw and Edward J. Valauskas studied the effectiveness of Dublin Core metadata to see if the use of metadata enhanced information retrieval in a suite of specific search engines, and the results suggested that metadata did not play a significant role in increasing the discovery of the resources.[26] In a later study, based on the statistical analysis on six search engines, Safari found a similar result, concluding that using Dublin Core metadata elements did not improve the retrieval ranking of the Web pages.[27] In other words, deployment of Dublin Core metadata elements in Web page resources did not affect retrievability of the resources, nor was it an impact factor for resource discovery on the Web. Safari ascribed the result to the lack of consensus on implementing Dublin Core, suggesting that the metadata schema was not widely accepted and used by search engines.[28]

In addition to the effectiveness study of metadata in discoverability of online resources, ratings of metadata elements is another aspect that metadata quality evaluation involves. Yen Bui and Jung-Ran Park conducted a research of the repository metadata for quality evaluation and found that five essential metadata elements, regardless of types of metadata schema, include Descriptor, Title, Subject, Type and Identifier.[29] In 2012, Phelps surveyed 236 national library Web sites and found that the five most common Dublin Core elements were Date, Title, Language, Creator, and Subject.[30] In a most recent study of metadata usage on three math and science digital repositories, Windnagel found that the frequently used Dublin Core elements were Identifier, Title, Description, and Contributor, which were also required or strongly recommended during implementation.[31] None of the mentioned research, however, built the connection between the rankings of metadata and discoverability by Internet search engines.

RESEARCH QUESTIONS

Although discussions on metadata quality and evaluation criteria have lasted for almost two decades, few evaluations have been conducted. Instead of evaluating metadata of digital collections, the research primarily evaluated effectiveness of metadata on Web page resource discovery by search engines. The research of metadata ranking mainly considered metadata usage in digital collection. None of them established a meaningful ranking between metadata elements and discoverability.

In order to find out how implemented metadata elements facilitate discovery of digital collections by search engines, this study employed a method to compare the values of metadata elements against keywords used by patrons in Internet search engines. By doing this, the study was able to evaluate the effectiveness of metadata by examining which implemented metadata elements have the highest keyword matching rates.

Research Question 1: Which metadata elements in the digital collection contain values that match the keywords patrons use when searching in Internet search engines?

Research Question 2: What is the keyword matching rate of each metadata element? Which metadata element is the most effective in facilitating discovery of digital collections by the Internet search engines?

METHODOLOGY

[insert institute's name] Libraries use DSpace as the main platform for the digital and institutional repository. The repository contained 46 digital collections and a total of 29,705 digital items as of July 31, 2014 according to the built-in statistics tool. The majority of the digital repository was electronic theses and dissertations, digitized documents and images, authenticated architecture materials, government documents collections, and research papers from other institutions on campus. Except for the one collection developed specifically for the Architecture Library using a combination of VRA Core and Dublin Core, the rest of digital collections were all implemented with Dublin Core metadata schema only. This study would not discuss selections of or have any preferences on metadata schemas. Values of metadata elements in all associated items, regardless of types of metadata schemas being used, were examined and compared with searching keywords.

The researcher utilized Google Analytics in this study for the purpose of data retrieving, organic search sorting, and keyword analyzing. Organic search is keyword searching behavior performed through all unpaid search engine mediums.[32] Google Analytics records traffics from

the major search engines, what search terms were used, and links to the page that the search term pulled up. Google Analytics only records this information if the search resulted in users visiting the repository.

Differentiation of search engines was not discussed in this study, thus organic searching keywords brought by different Internet search engines, including Google, Yahoo, Bing, etc., were treated equally. The selected time range for data extraction from the digital repository was between August 1, 2013, and July 31, 2014. [insert institute's name] Libraries completed content migration to DSpace in July 2013, the selected time range provided a whole year's data for the research.

According to Figure 1, during the selected time period the total count of visit sessions to the digital repository was 73,341 and the organic search shared 59% of the total, which was 43,016 sessions. The data demonstrated that search engine traffic, in the selected time frame, was the main traffic source of the digital repository, which contained a representative sample pool for the study. Traffic referred by Web domains were categorized as referral traffic, including Google Scholar, and these sent 20,763 visits (28%) to the digital repository in the selected time frames. Typing a URL into the browsers or visiting through bookmark tools were classified as direct traffic, which sent 8,815 visits (12%). Social Medias such as Facebook and Twitter made a 1% contributions to the traffic sources.

[Insert Figure 1]

The total number of keywords that were searched during the selected time period and successfully helped the patron navigate to the digital repository was 22,559, all of which were associated with one or multiple landing page URLs. Keywords' association with landing page URLs indicated that the keyword searched in Internet search engines all resulted in landing to the digital repository. Google Analytics sorted 22,559 keywords by the frequency of searching in default, but Google Analytics provided other means for sorting, such as by the alphabetical order of keywords.

The ideal sample size for a population of 22,559 should be 378, with a 95% Confidence Level and a ±5% Margin of Error. Table 1 summarizes the basic parameters of the research.

[Insert Table 1]

In order to guarantee samples were selected at random, the researcher used Random Integer Generator[33] to generate 378 random numbers ranging from 1 to 22,559, then sorted the keywords in Google Analytics by alphabetical order and chose each numbered keyword by using the random numbers. The researcher followed the associated URLs with those keywords to each

landing page in the digital repository and compared the keywords against the values in all metadata elements of that item.

When a valid string in the keyword matched a text string in metadata elements the metadata element was recorded in an Excel table as matching searching keyword once. A valid string here was defined as a readable and meaningful word and phrase, regardless of languages. For example, the 20,530th keyword (the number was retrieved from the 378 random numbers) in this study was "the space syntax methodology 翻译" where "syntax methodology" and "翻译" were both regarded as valid words and compared against the metadata.

Although DSpace provides full-text indexing for PDF files[34], this research did not include or analyze the text content in associated PDF files. Moreover, sub-fields of qualified metadata element were regarded as a main element. For example, an abstract of a digital item in the metadata element dc.description.abstract received examination and comparison against the keywords, and was recorded in the result as the dc.description.

After visiting each landing page and completing metadata-keyword comparison, the researcher recorded the results in an Excel table for further data sorting, calculating, and figure drawing. By doing this, the researcher was able to identify which metadata elements contained values that were matched with keywords and to calculate the matching rate of each matched metadata element.

RESULTS

The researcher visited each of 378 keywords in Google Analytics and associated landing pages in the digital repository. Of the samples, 377 were valid keywords associated with valid landing pages, meaning that these keywords, searched by the Internet search engines, matched values of certain metadata elements in the digital repository and resulted in visits to the associated landing pages. One keyword, however, was associated with an invalid URL directing to a page containing "Page Not Found" information. It is possible that this particular digital item was deleted from the digital repository after the visit was recorded by Google Analytics. However, one invalid sample in a sample size of 378 for the population 22,559 did not significantly affect the margin of error. The result of the research was still sound. Table 2 summarizes the corrected information of the research.

[Insert Table 2]

Examining the 377 keyword samples, the research identified that in total six Dublin Core metadata elements and two VRA Core metadata elements contained values that matched with the

keywords from the Internet search engines. As shown in Table 3, six Dublin Core metadata elements were Title, Creator, Contributor, Description, Subject, and Identifier. Two VRA Core metadata elements were Agent and Location.

[Insert Table 3]

Some of the metadata elements matched with the keywords by the Internet search engine more frequently than others. Some keywords found the matched values in only one metadata element while some keywords matched with multiple ones. Regardless of the overlap matching, Title was the metadata element which had the most keyword matches. Out of the sample, 279 (74.01%) keywords matched the values of the dc.title (See Table 3). The findings indicate that most keywords searched through the Internet search engines discovered the matched values of Title element in the digital repository and resulted in visits to the relevant items.

As shown in Table 3, the metadata element bringing the second most landing visits from the Internet search engines was Description, which received a 208 (55.17%) keyword matches out of 377. Subject element matched its values with 79 keywords (20.95%).

Not many author name related keywords brought landing visits from the Internet search engine according to Table 3. The Creator element only had 13 matches and Contributor element only had five. Identifier element received two keyword matches, but the researcher noticed that these two values contained citation information with author names, that indeed matched the keywords. Two VRA Core elements had only three keyword matches in total, two of which, however, had overlapping matches with the values in dc.title.

Figure 2 visually illustrates how frequently each metadata matched values with the keywords from Internet search engines. As seen in Figure 2, three metadata elements, dc.title, dc.description, and dc.subject contained the most matched values with keywords from Internet search engines, and caused most landing visits to the digital repository. The finding indicate that these three Dublin Core metadata elements played significant roles in facilitating discovery of digital collections through the Internet search engine.

[Insert Figure 2]

In order for the researcher to draw a more precise conclusion on the findings, Table 4 presents how each metadata element performed in the study by considering parameters and deviations. Regarding Dublin Core's Title element, for example, the researcher is now 95% certain that among the total search engine traffic to [insert institute's name] Libraries' digital repository, between 69.62% (74.01% - 4.39%) and 78.40% (74.01% + 4.39%) of keywords from

Internet search engines discovered the matched values in the metadata element dc.title in the digital repository and resulted in landing visits.

[Insert Table 4]

Similar conclusions can also be made for the rest of metadata elements. With a confidence interval of 95%, between 50.19% and 60.15% of keywords from Internet search engines discovered matched values in the metadata element dc.descirption; and between 16.88% and 25.02% of keywords did so in dc.subject. Table 4 also shows that the range for dc.creator was 3.45% (±1.83%) and for dc.contributor was 1.33% (±1.15%). The rest of the three had a small enough matching rate that should be confident enough, so that a confidence interval was not applicable.

The analysis between metadata elements and keywords so far has not removed the overlapping matches from the calculations. The researcher also found that some keywords had matched values in multiple metadata elements, while some keywords matched only one metadata element or even none (See Table 5).

[Insert Table 5]

According to Table 5, there were ten keywords which did not match values of any metadata elements. With the question of how these searches resulted in landing visits, the researcher found that they were actually referred from Google Scholar and those keywords contained a combination of digits and letters, such as the 17,461th keyword "related:zsti6noyraqj:scholar.google.com/." These visits referred by Google Scholar ideally should be regarded as Referral Traffic[35] defined in Google Analytics. However, Google Analytics mistakenly regarded them as organic search engine traffic.

More than half of keywords (51.72%, 195 keywords) found matched values at least in one metadata element, 123 keywords (32.63%) in two metadata elements, and 48 of them in three. Table 5 also shows that the keyword from the Internet search engine matched at most four metadata element values in the digital repository. In order to find out how well each metadata element performed in keyword matching, either by the element itself or combining with others, this research looked into more detailed analysis on the findings.

Table 6 shows that among the 195 keywords that discovered matched values in only one metadata element, the Dublin Core element dc.title had the most matched keywords with 112, 57.44% of 195 and 29.71% of 377. The finding evidenced that the Title element of Dublin Core by itself, regardless of the overlapping matches in Table 3, played a critical role in attracting keywords from Internet search engines. The Dublin Core element dc.description received the

second most matched keywords, with 65 (33.33%; 17.24%). These two Dublin Core elements were the most effective metadata elements in facilitating discovery when removing the influence of overlapping match.

Of particular note is the Dublin Core Subject element dc.subject. Comparing to Table 3, where dc.subject had 79 matched keywords (20.95% of 377) due to overlapping matches, the same metadata element only matched three keywords (1.54%; 0.80%) by itself. For complete information of all metadata elements, see Table 6.

[Insert Table 6]

Figure 3 presents a visualized chart for an overview of how each metadata element independently performed with matched keywords. Comparing with Figure 2, the difference demonstrated that when taking the overlapping match issue out from analysis, metadata elements dc.title and dc.description were still of significance in facilitating discovery of digital items through Internet search engines.

[Insert Figure 3]

Table 7 presents information of two metadata elements in combination finding matched keywords. Metadata elements dc.title, dc.description, and dc.subject were still key factors in facilitating discovery because these three metadata elements, combining with each other, had the most matched keywords from Internet search engines.

[Insert Table 7]

Table 8 displays information of three metadata elements in combination finding matched keywords. The data shows that metadata elements dc.title, dc.description, and dc.subject, combining together, had the most matched keywords from Internet search engines.

[Insert Table 8]

Table 9 records the only one keyword that discovered matched values from four metadata elements. And not surprisingly, metadata elements dc.title, dc.description, and dc.subject, again, were all in here.

[Insert Table 9]

DISCUSSION

The research results show that the randomly selected keywords, searching from Internet search engines, discover the matched values in six Dublin Core metadata elements and two VRA Core metadata elements in the digital repository. The six Dublin Core metadata elements are Title, Description, Subject, Creator, Contributor, and Citation; and VRA Core metadata elements are Agent and Location. Among these eight metadata elements, Dublin Core elements Title, Description, and Subject contain values that have most matched keywords and result in most landing visits to the digital repository. The finding indicates that these three metadata elements play the most significant roles in facilitating discovery of digital items by Internet search engines.

The Dublin Core metadata element dc.title in total has 279 matched keywords and a 74.01% matching rate. Considering the margin of error, it is 95% possible that among the total search engine traffics to the digital repository, between 69.62% and 78.40% of keywords from the Internet search engines search against the value of dc.title and result in landing visits. Similar conclusions can be drawn for another two Dublin Core metadata elements, Description and Subject. The range of keyword matching rate for dc.description is between 50.19% and 60.15%, and for dc.subject the range of matching rate is between 16.88% and 25.02%.

Dublin Core elements Creator and Contributor seem to play minor roles in facilitating discovery. In the same confidence level 95%, the range of matching rate for metadata element dc.creator is between 1.62% and 5.28%, and for dc.contributor is between 0.18% and 2.48%. Although the Dublin Core element Identifier has two matched keywords, but the two metadata values contain citation information and the matched keywords are the author names. Moreover, the matching rate is too small for a confidence interval, which means it is confident enough that keywords finding matched values in the citation element should be rare.

If removing the overlapping match issues from analysis, however, and considering only the independence matching rate for each metadata element, circumstances become slightly different. Metadata element dc.subject independently only has three matched keywords and the matching rate is (1.54%; 0.80%), lower than dc.creator (4.62%; 2.39%) and dc.contributor (2.56%; 1.33%) (See Figure 3). Granted, the Subject metadata element cannot be ignored because it still has up to 25.02% (See Figure 2), combining with other elements, of matched keywords from Internet search engines. The two Dublin Core metadata elements, dc.title and dc.description,

consistently have significantly higher matching rates than others, that are (57.44%; 29.71%) for dc.title and (33.33%; 17.24%) for dc.description.

CONCLUSION

The functionality of metadata for facilitating discovery of online resources has been a key issue in the discussion of metadata quality evaluations. A few early studies were conducted for testing the correlation between metadata of Web resources and Internet search engines, but these studies yield contrasting conclusions. Although few quality-ranking studies of metadata are found, no research is found in the literature regarding effectiveness of metadata between digital collections and Internet search engines.

In order to find out the answer to what metadata elements in digital collections are effectively facilitating the discovery of the resources through Internet keyword searching, and how much significance each metadata element play in this role, the researcher designed a method for the research. By retrieving data from [insert institute's name] Libraries' digital repository in DSpace, the researcher determined an ideal sample size for representing the whole population of organic search keywords and search engines traffics. The researcher then compared each randomly selected keyword against the associated digital item's metadata value, to determine the matching rate for metadata elements.

Based on the results, it can be concluded that three most significant metadata elements in enhancing discovery of digital repository items through Internet search engines are Title, Description, and Subject, which in this case are Dublin Core metadata elements dc.title, dc.description, and dc.subject. Research results remind librarians that when implementing metadata with digital items, they should pay specific attention to these three metadata elements if the institution aims to attract more traffic from Internet search engines. For example, a metadata librarian can try to input more detailed and precise information in the Description fields, which will enhance the discoverability of the digital items on search engines. Moreover, the results can also lead to a re-classification within metadata schemas to group metadata elements based on their functionality. Based on the findings, for example, scholars can categorize the three Dublin Core metadata elements into a searchability or discoverability group. Re-classifying the metadata elements based on the functionality can help professionals better understand the function of each specific element while implementing metadata.

Academic libraries have invested much time and effort in developing digital collection and institutional repositories, having insight into how these digital assets are discovered by search engines are necessary and helpful. The findings of this research fills the blank of the effectiveness study of metadata's facilitating discovery of digital repository items through Internet search engines, and contributes to the literature of metadata quality and metadata evaluations by pointing out a new direction for relevant research. Further studies, however, are still needed to rectify the limitations of this research.

The data for the study is retrieved only from one digital repository. Further research needs to be conducted on more digital repositories to see if a similar conclusion can be reached. A more specific analysis on targeted metadata elements can also be conducted in the future research to reach a more precise finding. For example, researchers can design a specific study to include designated keywords and metadata element values, so as to test searching keywords and to observe if the keywords search against and navigate to the targeted metadata elements. A comparison study can also be carried out to test the correlation of same values in different metadata elements. In doing so, the researcher can observe if retrievability is affected when designated metadata value is included in one metadata element instead of the other. Both non-PDF digital items and PDF-included digital items are needed in comparison study as well, to gauge the effects of PDF full-text indexing features provided by the digital repository system and the search engines.

NOTES

1. Ann Windnagel, "The Usage of Simple Dublin Core Metadata in Digital Math and Science Repositories," *Journal of Library Metadata* 14, no. 2 (2014): 77-102.
2. Diane I. Hillmann, "Metadata Quality: From Evaluation to Augmentation," *Cataloging & Classification Quarterly* 46, no. 1 (2008): 65-80; R. J. Robertson, "Metadata Quality: Implications for Library and Information Science Professionals," *Library Review* 54, no. 5 (2005): 295-300.
3. Matthew Beacom, *Crossing a Digital Divide: AACR2 and Unaddressed Problems of Networked Resources* (ERIC Clearinghouse, 2000); Ann Huthwaite, "AACR2 and Its Place in the Digital World: Near-Term Solutions and Long-Term Direction" (Nov. 2000), available online at http://files.eric.ed.gov/fulltext/ED454865.pdf [accessed 15 August 2014]; Amy K. Weiss and Timothy V. Carstens. "The Year's Work in Cataloging, 1999," *Library Resources & Technical Services* 45, no. 1 (2001): 47-58; Mehdi Safari, "Search Engines and Resource Discovery on the Web: Is Dublin Core an Impact Factor?," *Webology* 2, no. 2 (2005), available online at http://www.webology.org/2005/v2n2/a13.html [accessed 15 August 2014].
4. Safari, "Search Engines and Resource Discovery on the Web."
5. Jung-Ran Park, "Metadata Quality in Digital Repositories: A Survey of the Current State of the Art," *Cataloging & Classification Quarterly* 47, no. 3-4 (2009): 213-228.
6. F. O. Isinkaye, A. B. C. Robert, and B. A. Ojokoh, "An Evaluation of Metadata Integrity in Textual Documents," *Journal of Library Metadata* 12, no. 1 (2012): 1-14.
7. Windnagel, "The Usage of Simple Dublin Core Metadata in Digital Math and Science Repositories," 77-78.
8. Kathleen Burnett, Kwong Bor Ng, and Soyeon Park, "A Comparison of the Two Traditions of Metadata Development," *Journal of the American Society for Information Science* 50, no. 13 (1999): 1209-1217.

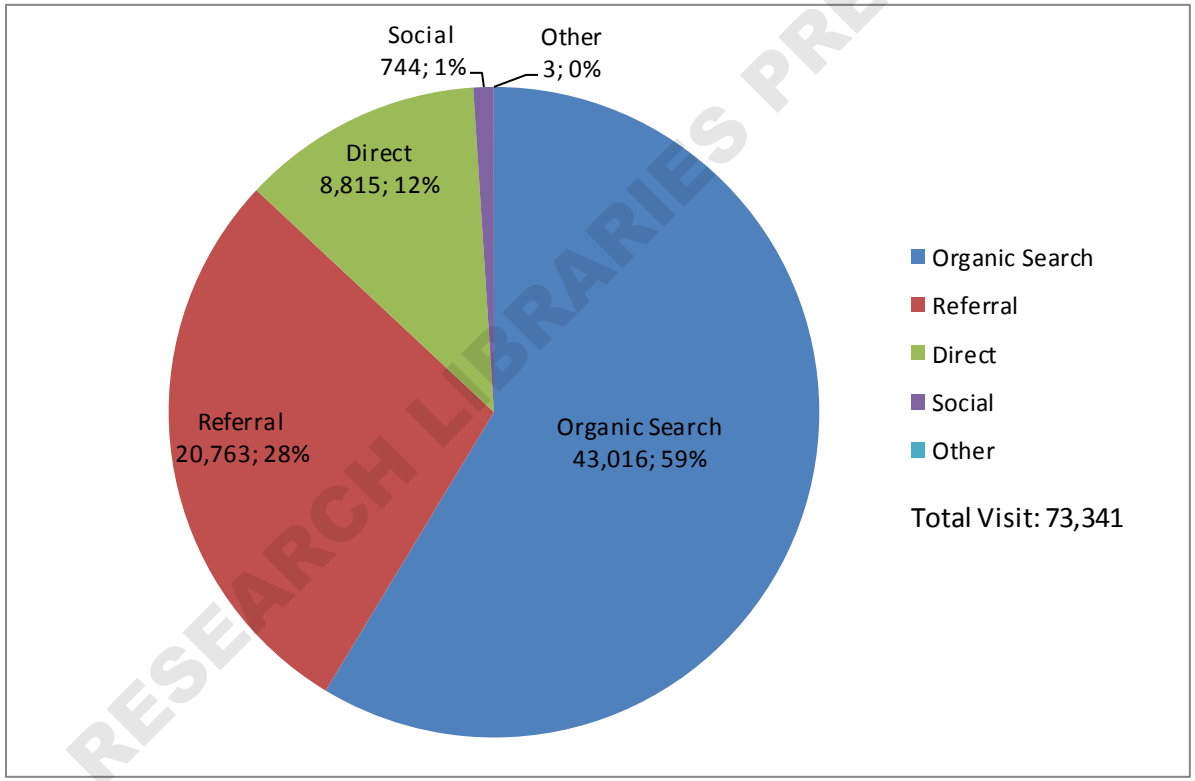
9. Kuang-Hwei Lee–Smeltzer, "Finding the Needle: Controlled Vocabularies, Resource Discovery, and Dublin Core," *Library Collections, Acquisitions, and Technical Services* 24, no. 2 (2000): 205-215.
10. National Information Standards Organization, "Understanding Metadata" (Bethesda, MD: NISO Press, 2004), available online at http://www.niso.org/publications/press/UnderstandingMetadata.pdf [accessed 15 August 2014].
11. Hillman, "Metadata Quality," 67.
12. William E. Moen, Erin L. Stewart, and Charles R. McClure. "The Role of Content Analysis in Evaluating Metadata for the U.S. Government Information Locator Service (GILS): Results from an Exploratory Study" (1997), available online at http://digital.library.unt.edu/ark:/67531/metadc36312/ [accessed 15 August 2014].
13. Thomas R. Bruce and Diane I. Hillmann, "The Continuum of Metadata Quality: Defining, Expressing, Exploiting," in Metadata in Practice, eds. D. Hillmann and E. L. Westbrooks (Chicago: American Library Association, 2004).
14. Park, "Metadata Quality in Digital Repositories," 215.
15. Sevim McCutcheon, "Keyword vs Controlled Vocabulary Searching: The One with the Most Tools Wins," *The Indexer* 27, no. 2 (2009): 62-65.
16. Tyler E. Phelps, "An Evaluation of Metadata and Dublin Core Use in Web-based Resources," *Libri* 62, no. 4 (2012): 326-335.
17. Warwick Cathro, "Metadata: An Overview" (1997), available online at http://www.nla.gov.au/openpublish/index.php/nlasp/article/view/1019/1289 [accessed 15 August 2014].
18. Stuart Weibel, Jean Godby, Eric Miller, and R. Daniel, "OCLC/NCSA Metadata Workshop Report" (1995), available online at http://ci.nii.ac.jp/naid/10011891419 [accessed 15 August 2014].

19. NISO, “Understanding Metadata.”
20. Ininkaye, Robert, and Ojokoh, “An Evaluation of Metadata Integrity in Textual Documents.”
21. Safari, “Search Engines and Resource Discovery on the Web.”
22. Ibid.
23. Thomas P. Turner and Lise Brackbill, “Rising to the Top: Evaluating the Use of the HTML Meta Tag to Improve Retrieval of World Wide Web Documents through Internet Search Engines,” *Library Resources & Technical Services* 42, no. 4 (1998): 258-271.
24. Jin Zhang and Alexandra Dimitroff, “Internet Search Engines’ Response to Metadata Dublin Core Implementation,” *Journal of Information Science* 30, no. 4 (2004): 310-320.
25. Ibid.
26. Robin Henshaw and Edward J. Valauskas, “Metadata as a Catalyst: Experiments with Metadata and Search Engines in the Internet Journal, First Monday,” *Libri* 51, no. 2 (2001): 86-101.
27. Safari, “Search Engines and Resource Discovery on the Web.”
28. Ibid.
29. Yen Bui and Jung-Ran Park, “An Assessment of Metadata Quality: A Case Study of the National Science Digital Library Metadata Repository” (2006), available online at http://www.academia.edu/download/31045967/bui_2006.pdf [accessed 15 August 2014].
30. Phelps, “An Evaluation of Metadata and Dublin Core Use in Web-based Resources.”
31. Windnagel, “The Usage of Simple Dublin Core Metadata in Digital Math and Science Repositories.”
32. Google, “Traffic Source Dimensions,” available online at https://support.google.com/analytics/answer/1033173?hl=en [accessed 15 August 2014].
33. Random Integer Generator, available online at http://www.random.org/integers/?num=378&min=1&max=22559&col=10&base=10&format=html&rnd=new [accessed 15 August 2014].

34. DuraSpace, “Configure Full Text Indexing,” available online at https://wiki.duraspace.org/display/DSPACE/Configure+full+text+indexing [accessed 15 August 2014].
35. Google, “Referral Traffic,” available online at https://support.google.com/analytics/answer/1247839?hl=en&ref_topic=1631856 [accessed 15 August 2014].

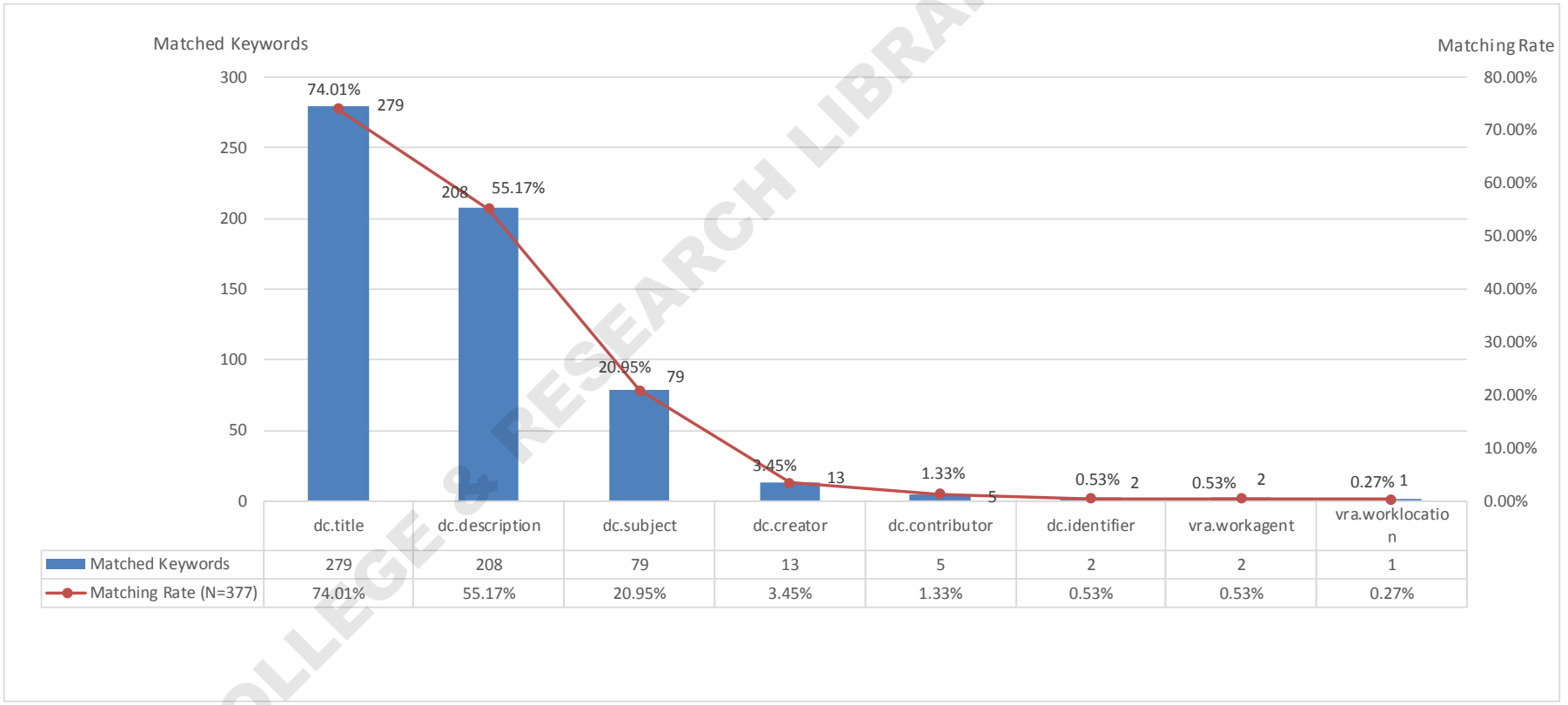
Organic Search	43,016	
Referral	20,763	
Direct	8,815	
Social	744	
Other	3	

FIGURE 1
Traffic Sources to Digital Repository (August 1, 2013 - July 31, 2014)



Metadata Elements	Matched Keywords	Matching Rate (N = 377)
dc.title	279	74.01%
dc.description	208	55.17%
dc.subject	79	20.95%
dc.creator	13	3.45%
dc.contributor	5	1.33%
dc.identifier	2	0.53%
vra.workagent	2	0.53%
vra.worklocation	1	0.27%

FIGURE 2
Frequency of Metadata Elements Discovering Matched Keywords (N = 377)



Metadata Elements	Matched Keywords	Matching Rate (n = 195)	Matching Rate (N = 377)
dc.title	112	57.44%	29.71%
dc.description	65	33.33%	17.24%
dc.creator	9	4.62%	2.39%
dc.contributor	5	2.56%	1.33%
dc.subject	3	1.54%	0.80%
vra.workagent	1	0.51%	0.27%
dc.identifier	0	0.00%	0.00%
vra.worklocation	0	0.00%	0.00%

FIGURE 3
1 Metadata Element & Matched Keyword (n = 95 ; N = 377)

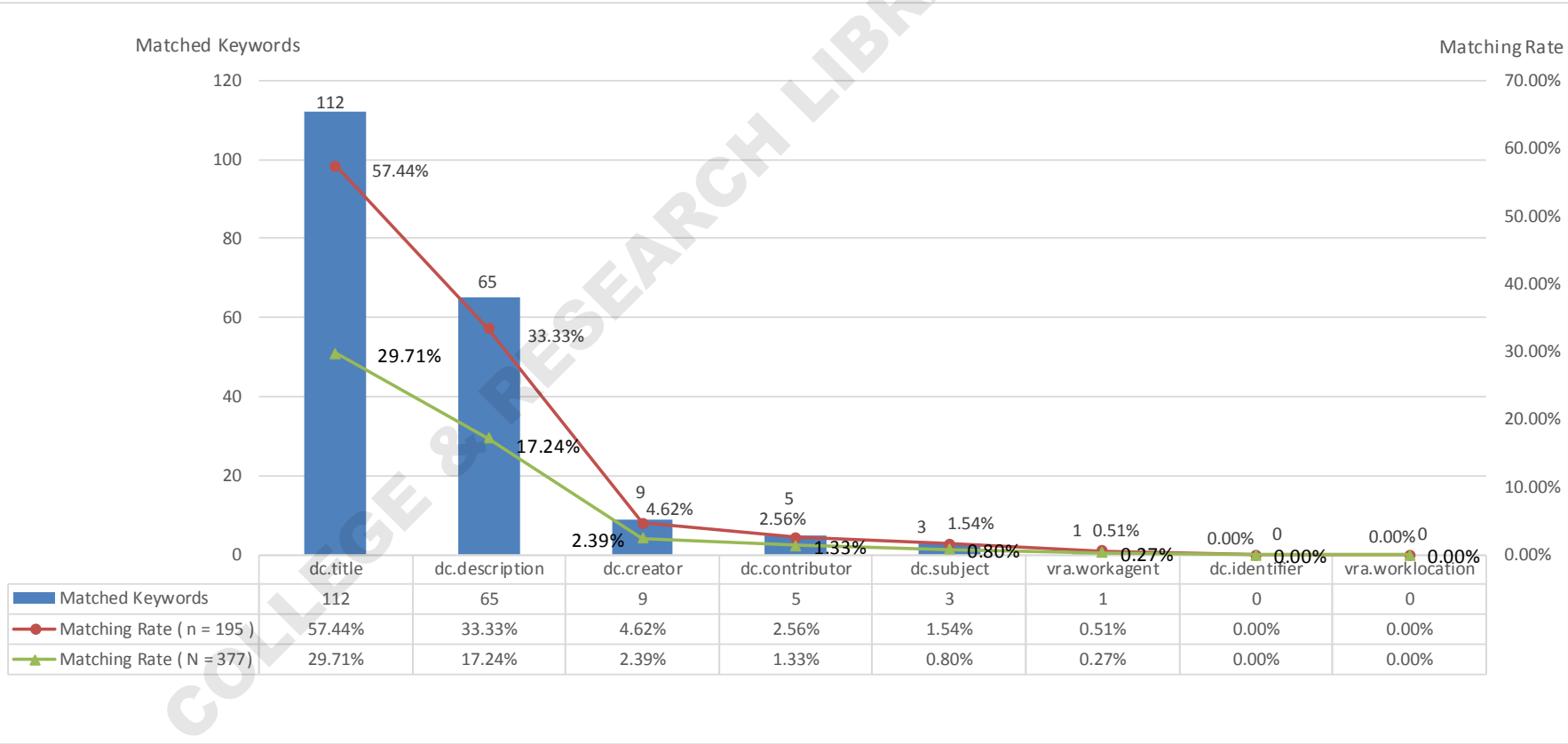


TABLE 1 Parameters of the Research		
Keyword Population	22,559	Population = 22,559
Confidence Level	95%	Z = 1.96
Margin of Error (Confidence Interval)	±5%	c = 0.05
Standard of Deviation	0.5	p = 0.5
Sample Size	377.74	Sample Size = 378

TABLE 2 Corrected Parameters of the Research	
Keyword Population	22,559
Sample Size	378
Valid Samples (Keywords with valid URLs)	N = 377
Invalid Samples (Keywords with invalid URLs)	1
Confidence Level	95%
Margin of Error (Confidence Interval)	±5%
Standard of Deviation	0.5

TABLE 3
Metadata Elements, Matched Keywords, & Matching Rate (N = 377)

Metadata Schema	Metadata Element	XML Element in Digital Repository	Matched Keywords	Matching Rate
Dublin Core	Title	dc.title	279	74.01%
	Creator	dc.creator	13	3.45%
	Contributor	dc.contributor	5	1.33%
	Description	dc.description	208	55.17%
	Subject	dc.subject	79	20.95%
	Identifier	dc.identifier	2	0.53%
VRA Core	Agent	vra.workagent	2	0.53%
	Location	vra.worklocation	1	0.27%

TABLE 4
Deviation of Each Metadata Element (N = 377; Confidence Level = 95%; Z = 1.96; Population = 22,559)

Metadata Element	Matched Keywords	Matching Rate	Margin of Error
dc.title	279	74.01%	±4.39%
dc.description	208	55.17%	±4.98%
dc.subject	79	20.95%	±4.07%
dc.creator	13	3.45%	±1.83%
dc.contributor	5	1.33%	±1.15%
dc.identifier	2	0.53%	N/A
vra.workagent	2	0.53%	N/A
vra.worklocation	1	0.27%	N/A

TABLE 5 Number of Metadata Elements & Matched Keywords (N = 377)		
Number of Metadata Elements	Matched Keywords	Matching Rate
0	10	2.65%
1	195	51.72%
2	123	32.63%
3	48	12.73%
4	1	0.27%

TABLE 6 1 Metadata Element & Matched Keywords (n = 195 ; N = 377)			
Metadata Element	Matched Keywords	Matching Rate (n=195)	Matching Rate (N=377)
dc.title	112	57.44%	29.71%
dc.creator	9	4.62%	2.39%
dc.contributor	5	2.56%	1.33%
dc.description	65	33.33%	17.24%
dc.subject	3	1.54%	0.80%
dc.identifier	0	0.00%	0.00%
vra.workagent	1	0.51%	0.27%
vra.worklocation	0	0.00%	0.00%

TABLE 7
2 Metadata Elements & Matched Keywords (n = 123 ; N = 377)

Metadata Element		Metadata Element	Matched Keywords	Matching Rate (n=123)	Matching Rate (N=377)
dc.title	And	dc.description	89	72.36%	23.61%
dc.title		dc.subject	26	21.14%	6.90%
dc.title		dc.creator	1	0.81%	0.27%
dc.title		vra.workagent	1	0.81%	0.27%
dc.title		vra.worklocation	1	0.81%	0.27%
dc.description		dc.subject	5	4.07%	1.33%

TABLE 8
3 Metadata Elements & Matched Keywords (n = 48 ; N = 377)

Metadata Element	And	Metadata Element	And	Metadata Element	Matched Keywords	Matching Rate (n=48)	Matching Rate (N=377)
dc.title		dc.description		dc.subject	44	91.66%	11.67%
dc.title		dc.description		dc.identifier	2	4.17%	0.53%
dc.title		dc.description		dc.creator	2	4.17%	0.53%

TABLE 9
4 Metadata Elements & Matched Keywords (n = 1 ; N = 377)

Metadata Element	And	Metadata Element	And	Metadata Element	And	Metadata Element	Matched Keywords	Matching Rate (n=1)	Matching Rate (N=377)
dc.title		dc.description		dc.subject		dc.creator	1	100.00%	0.27%