



*Evidence Summary*

**Title, Description, and Subject are the Most Important Metadata Fields for Keyword Discoverability**

**A Review of:**

Yang, L. (2016). Metadata effectiveness in internet discovery: An analysis of digital collection metadata elements and internet search engine keywords. *College & Research Libraries*, 77(1), 7-19.  
<http://doi.org/10.5860/crl.77.1.7>

**Reviewed by:**

Laura Costello  
Head of Research & Emerging Technologies  
Stony Brook University Libraries  
Stony Brook, New York, United States of America  
Email: [laura.costello@stonybrook.edu](mailto:laura.costello@stonybrook.edu)

**Received:** 1 June 2016

**Accepted:** 15 July 2016

© 2016 Costello. This is an Open Access article distributed under the terms of the Creative Commons-Attribution-Noncommercial-Share Alike License 4.0 International (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly attributed, not used for commercial purposes, and, if transformed, the resulting work is redistributed under the same or similar license to this one.

---

**Abstract**

**Objective** – To determine which metadata elements best facilitate discovery of digital collections.

**Design** – Case study.

**Setting** – A public research university serving over 32,000 graduate and undergraduate students in the Southwestern United States of America.

**Subjects** – A sample of 22,559 keyword searches leading to the institution's digital repository between August 1, 2013, and July 31, 2014.

**Methods** – The author used Google Analytics to analyze 73,341 visits to the institution's digital repository. He determined that 22,559 of these visits were due to keyword searches. Using Random Integer Generator, the author identified a random sample of 378 keyword searches. The author then matched the keywords with the Dublin Core and VRA Core metadata elements on the landing page in the digital repository to determine which metadata field had drawn the keyword searcher to that particular page. Many of these keywords matched to more than one metadata field, so the author also analyzed the metadata elements that generated unique keyword hits and those fields that were frequently matched together.

**Main Results** – Title was the most matched metadata field with 279 matched keywords from searches. Description and Subject were also significant fields with 208 and 79 matches respectively. Slightly more than half of the results, 195 keywords, matched the institutional repository in one field only. Both Title and Description had significant match rates both independently and in conjunction with other elements, but Subject keywords were the sole match in only three of the sampled cases.

**Conclusion** – The Dublin Core elements of Title, Description, and Subject were the most frequently matched fields in keyword searches. Academic librarians should focus on these elements when creating records in digital repositories to optimize traffic to their site from search engines.

### Commentary

This study examines common digital repository metadata fields by looking critically at successful keyword searches and provides context for the way these records are discovered organically through search engine traffic. Though both of these topics have been explored independently, the latter largely outside of library literature, the study represents a unique illumination of library metadata through the lens of general searching. A few studies have examined the frequency of Dublin Core elements on websites (Phelps, 2012; Windnagel, 2014), though this study is unique in its consideration of these elements through external search engines. Though projects like linked open data and current metadata schema development deeply consider the impact of digital searching, the results of this study could potentially lead to search-oriented workflow optimization in existing collections. The study's focus on keywords for searching is particularly helpful for libraries struggling to make in-house digital collections more visible in discovery layers and through organic searches from outside the library.

The author chose an appropriate sample size for a 95% confidence level and a  $\pm 5\%$  margin of error, and samples were selected randomly over the course of one year of data. The sample selected seems likely to be representative of the types of searches that are regularly performed by users when accessing the digital repository. The major limitation of this study is that it examines only one digital repository. More research is needed to determine whether the results are generalizable to other repositories with different collections.

Broadening this type of research to other collections is particularly important for studying search because much of the strategy and success of search practice is unique to the file type and format type of the material being searched. Though this study focused on a large digital repository of 29,705 items and included many of the common file formats and types found in digital repositories, such as digitized images and text, dissertations, and research papers, there is much to gain from testing the results against other collections.

Digital libraries have struggled with crafting metadata that accommodates and supports searches conducted within library catalogues and resources while providing enough information for non-library search engines. This study highlights the essential points of metadata creation from the perspective of outside searching but has the potential to reflect back on the way libraries internally evaluate appropriate and essential metadata for digital materials. As library searching becomes more keyword-based, it will be important to continue to study the way keyword searches interact with digital metadata.

### References

Phelps, T. E. (2012). An evaluation of metadata and Dublin Core use in web-based resources. *Libri: International Journal of Libraries & Information Services*, 62(4), 326-335. <http://doi.org/10.1515/libri-2012-0025>

Windnagel, A. (2014). The usage of Simple Dublin Core metadata in digital math and science repositories. *Journal of Library Metadata*, 14(2), 77-102.  
<http://doi.org/10.1080/19386389.2014.909677>