INTRODUCTION

Image Resource Findability on the World Wide Web is still very much a land-grab. For the Semantic Web to become a reality online businesses and individuals have to get their hands dirty and also come face-to-face with the realization that search engine plans are increasingly becoming the go-to-boo for information resource retrieval. “Increasingly, students use Web search engines such as Google to locate information resources rather than seek out library online catalogs or databases of scholarly journal articles” (Lippincott 2013). This puts the search engine giant in a unique position to dictate how the future of search will work on the Web – and therefore, your organization’s future presence (or lack thereof) on the Web.

There are no guardians of the Web, ensuring structured data is uniformly applied to all records with equal attention and care and there is no standard, mandated requirement for records on the Web to provide context for image resource findability. Most search engines do not crawl embedded XMP data or the invisible Web, leaving text near images, file names or text in the alt-text in html markup as the only context for image resources. The search algorithms for image retrieval are subject to change frequently (Kritzinger 2013) and additionally, social media sites and organizations strip embedded data from images (Embedded Metadata Manifesto 2014). Embedded metadata provides context and provenance for image resources.

RESEARCH METHODS

The following research question informed this project: What are the types and quality of structured data, XMP, and metadata records available for image resources appearing on the website?

Utilizing the Structured Data Linter Tool and Phil Harvey’s ExifTool, information was gathered to quantify these research questions. Image records on the Gateway to Oklahoma History’s website were investigated for the types, quality and quantity of embedded metadata and structured data.

RESULTS

The Gateway to Oklahoma History’s Website has a wealth of structured data and metadata pertaining to its digital records. Search queries utilizing structured data markup tags and/or embedded metadata yielded relevant and accurate results during a normal web search, but did not yield relevant and/or accurate image resources during an image search. Descriptive filenames were not used for image resources, which is an important part of image retrieval through search engines.

- Adding Schema.org tags to the on-page markup, to accompany the structured data already present is another area for improvement. An interesting finding from this research was that embedded metadata was only found on the largest, original version of the image resource, and never on smaller derivative images.
- Structured data included in the on-page mark-up included Open Graph Protocol and Dublin Core. IPTC was the another area for improvement.
- Adding Schema.org tags to the on-page markup, to accompany the structured data already present is another area for improvement. An interesting finding from this research was that embedded metadata was only found on the largest, original version of the image resource, and never on smaller derivative images.

CONCLUSION

The results and methodology for this research can help GLAM institutions (Galleries, Libraries, Archives & Museums) by bringing awareness to the state of structured data and image resource findability for cultural heritage institutions on the Web. GLAMs must be active in the SEO space, support machine-readable language in the markup of their sites, and utilize Schema.org vocabularies and descriptive filenames for relevancy in search engine results.

The Digital Library Federation, which is a program of the Council on Library and Information Resources, concludes that “Getting found means repository objects must be included in the indexes of major search engines because most students and faculty now begin their research with Internet search engines. Digital repositories created by libraries will be largely invisible to users if their contents are not indexed in these search engines” (Digital Library Foundation 2014).

REFERENCES


ACKNOWLEDGEMENTS

I would like to thank The University of Oklahoma, who offered resources to make this project possible. The Gateway to Oklahoma History Center also allows me to work on such a wonderful digital project for the Gateway to Oklahoma History during my study.