# Search Engine Optimization for Digital Repositories

Kenning Arlitsch, Associate Director for IT Services, University of Utah

Patrick OBrien, Managing Partner, RevX Corp ([www.RevXCorp.com](http://www.RevXCorp.com))

### Description

Surveys conducted by the University of Utah across numerous libraries and archives have revealed a disturbing reality: the number of digital objects successfully harvested and indexed by search engines from library digital repositories is abysmally low. The use of the scholarly and lay content in these databases is predicated on visibility in Internet search engines, and libraries have spent millions over the past decade creating repositories whose objects are being harvested and indexed only minimally. The reasons for the poor showings in Internet search engines are complex, and have both technical and administrative components.

Among the reasons why digital repositories may perform poorly in search engines:

1. Web servers may be configured incorrectly, and may lack sufficient speed performance;
2. Repository software may be designed or configured in a way that is difficult for crawlers to navigate;
3. Metadata are often not unique or structured as recognizable taxonomies;
4. Some search engines, such as Google Scholar, prefer schemas other than Dublin Core;
5. Search engine policies change, and some search engines are not supporting commonly accepted standards such as OAI-PMH.

The problem lies less with search engines than with the content that they are trying to harvest and index, but improvements can be made to the way the content is presented so that search engines can parse, organize, and serve more relevant results. The search engine market is fluid and intensely competitive. While Google retains the majority of direct search engine traffic, Bing is making progress quickly, and social media engines are changing the face of search itself, putting more emphasis on content that is popular and frequently refreshed. These changes will further affect the visibility of the content in library websites and their digital repositories.

The Marriott Library at the University of Utah has been working with RevX Corporation to develop a program that coordinates several layers of the library organizations, including IT, cataloging and metadata, marketing and publicity, and administration. Each of these departments plays a role in improving the reach and visibility of library websites and digital repositories. The program is being designed to develop actionable recommendations within a framework library staff can identify with and provide guidance on, and it uses data to help them communicate the value proposition of digital libraries.

The program started with a pilot in May 2010 that sought to increase the number of digital objects in the Google search engine and develop internal library staff skills to maintain and improve the program. The efforts have produced results in key areas.



Only 2% of the library’s 3,000+ EAD finding aids were included in Google’s general index in April 2010. As of January 2011 the number has increased to 69% and continues to climb.

The percentage of the library’s 145,000+ digital objects indexed by Google has increased from 12% in July 2010 to over 46% in January 2011. We have also achieved a 93% Google index ratio for a single digital collection with more then 500 URLs.



Substantial increases in the number of digital collection page views are evident since the initial pilot was implemented in May 2010.