

The importance of “marketing” digital collections: including a case study from Harvard’s Open Collections Program

Christine McCarthy Madsen

Introduction

An essential but easily overlooked step in the creation of online digital collections is making sure that people—potential users—know they sites exist. Making a digital collection publically available means that your potential user base—your potential patron pool—is the whole world. So what is being done to let that (significantly) wider audience know that your collection exists? Academic libraries have ways of letting patrons know what paper based collections they have and when new titles of interest arrive; and they have outreach programmes to let incoming students and faculty know what services they provide. But there seems to be a disconnect when it comes to online digital collections and often a “build it and they will come” attitude prevails.

This paper is presented in three parts. The first will introduce a method—inlink analysis—that can be used to identify and understand the current and potential audiences for digital collections. Part two will present some evidence from a case at Harvard University’s Open Collections Program (<http://ocp.hul.harvard.edu>) that marketing efforts have a significant impact on the number of users of digital collections. The third part will look more broadly at marketing in libraries; at what exactly academic libraries are and should be marketing.

Inlink Analysis

Inlinks are defined simply as any link to a web site. If four different web sites link to Web site A, it is said to have 4 inlinks. Inlink analysis is simply counting and analyzing the links to a given web site. Web log analysis and tools like Google Analytics can provide better data about your own web site, but link analysis allows you to get data about sites for which you do not have access. In this way inlink analysis tools can help you compare your site to others—which can be both an interesting exercise in understanding your audience and an important tool for sharing information with funders.

In March of 2008, 289 humanities-based digitization project sites were selected for inlink analysis.¹ The goal of this project was three fold:

1. To find out which of the initial 289 sites listed was the most linked to;
2. To determine if any visible pattern or perceptible social network emerged from the linking patterns between these sites; and

¹ ‘Digitization projects’ refers here to projects that involve the digital reformatting of texts, images, moving image, or audio into some sort of digital format.

3. To evaluate link analysis as a means to determine the relative ‘authority’ of these library digitization project sites.

This project relied on the basic assumption that a link from page A to page B is a recommendation of page B by the author of page A. It is, of course, not difficult to find exceptions to this, but for this purposes of this project it was generally accepted as valid—or at least as valid as in traditional citation analysis and bibliometrics. The success of citation analysis recognizes that there is a chance that someone has cited an article as an example of “what not to do,” but relies on the assumption that if an article is very heavily cited, it can be considered successful or authoritative.

The 289 sites were chosen based on an existing knowledge of the field, combined with targeted searches on project domains. All project links were verified as digitization project sites (as defined above) rather than just sites that described primary resources or library collections. Sites that contained entire works (books, manuscripts, etc.) were favoured over those that contained only a few page images.

Using a combination of LexiURL Searcher² and Yahoo! Site Explorer³, the number of inlinks to each of these sites was gathered and sorted in descending order. These simple counts of inlinks reveal at least one interesting pattern and unexpected result. Looking at the histograms for the top fifty and top one hundred sites (Appendix A), it is clear that the number sites containing numerous inlinks drops off sharply. The fourth most linked-to site contains only 16% of the inlinks of the most linked to site, while the tenth most linked to site contains only 4%. It seems that very few sites dominate this domain.

Cross referencing the inlink ranking of these sites with their usage rankings by the Alex Search Engine⁴ (Appendix B)—although not a precise match—does support that Project Gutenberg and Library of Congress American Memory (the two most linked-to sites) are also comparatively heavily used. But there is little we can know about how these sites gained such sizeable audiences. The Library of Congress has made significant efforts to connect directly with teachers through their Learning Page (<http://memory.loc.gov/learn/>) and invested in printed promotional materials such as brochures and posters, but one can only infer how Project Gutenberg came to be so heavily cited.

One of the interesting—and serendipitous—findings of this project was that two of the sites that I worked on for several years at Harvard have considerably different numbers of inlinks. In July of 2009, *Women Working* (<http://ocp.hul.harvard.edu/ww>) had 3,196 inlinks and on the same day *Immigration to the US*

² LexiURL and LexiURL Searcher: <http://lexiurl.wlv.ac.uk/>

³ Yahoo! Site Explorer: <https://siteexplorer.search.yahoo.com>

⁴ Alexa Search Engine, from Alexa the Web Information Company: <http://www.alexa.org/>

(<http://ocp.hul.harvard.edu/immigration>) had 844. The rest of this paper will be a discussion of how I think that came to be and why I think it is significant.

The Harvard Open Collections Program Case Study

The Marketing Campaign

The Open Collections Program was established in 2002 with a grant from the William and Flora Hewlett Foundation. There were two initial goals of the program:

- “to increase the availability and use of historical resources from Harvard's libraries, archives and museums for teaching, learning, and research”
- “to offer a new model for digital collections that will benefit students and teachers around the world”⁵

The program's first collection is called *Women Working* and is about the role of women in the US economy between 1700 and 1930. The collection contains over 500,000 pages of digitized primary resources⁶ including 7,500 pages of manuscripts, 3,500 books and pamphlets and 1,200 photographs

The program was also built around three main principles of quality:

- Selection Standards - Create comprehensive, topic-based digital collections by carefully selecting topics, and materials;
- Production Standards - Create digital surrogates that are both faithful to the original publications and of such high quality that there will be no need for re-digitization by other institutions;
- Access Standards - Provide easy online access to digital collections within the Harvard community and around the world.⁷

Within the last principle—access standards—we strongly believed that there existed a commitment not just to making the materials freely available online, but to making sure people knew these collections existed. With a grant renewal in 2004 we embarked on an outreach campaign to make people aware of the newly released but still not complete *Women Working* collection. This included:

- Hiring an ‘outreach and evaluation coordinator.’
- An email announcement campaign targeted at faculty, teachers, researchers, and librarians in relevant subject areas.

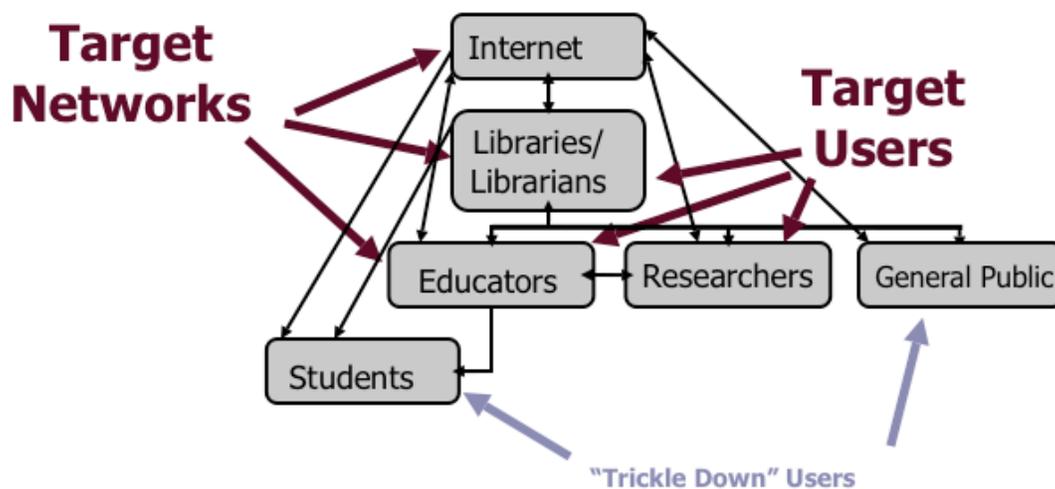
⁵ This text was on the initial home page of the Open Collections Program. It was unfortunately taken off and is not archived publicly.

⁶ Statistics from the homepage of *Women Working*: <http://ocp.hul.harvard.edu/ww>

⁷ Ibid.

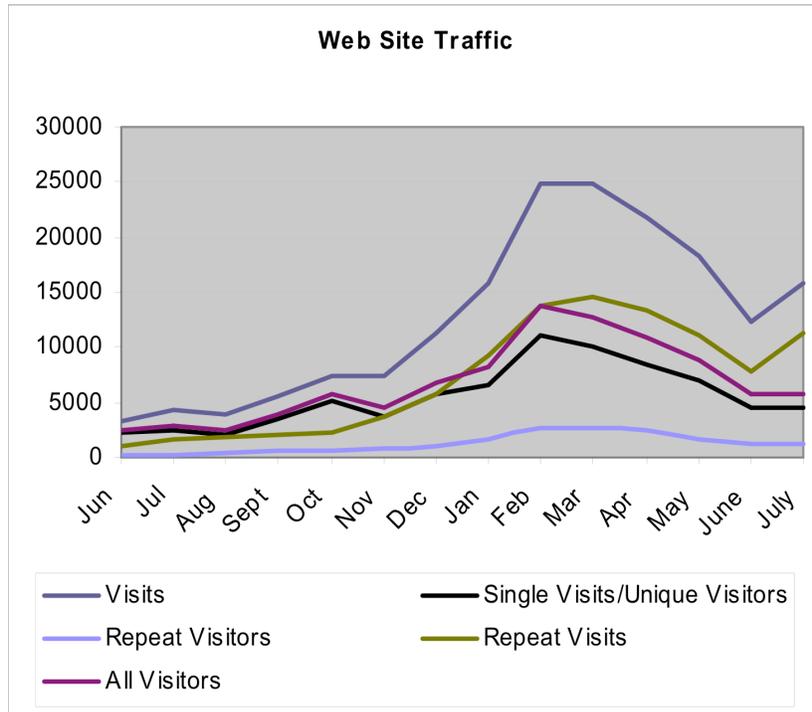
- Using inlink analysis and basic searching to identify course syllabi and descriptions to compile a list of relevant courses, instructors, and institutions and contacting individual instructors to engage them in using materials.
- Contacting the creators of similar sites to ask if they would link to our site.
- A mailing of printed brochures to over 2,000 librarians in relevant subject areas with information included on how they could receive more brochures for their library.
- A monthly email newsletter letting users know what was new to the collections and including small case studies of our users.
- Press releases to media outlets.
- Full-page advertisements in conference publications and relevant journals.

Target users and networks of users were identified as a key part of the outreach plan:

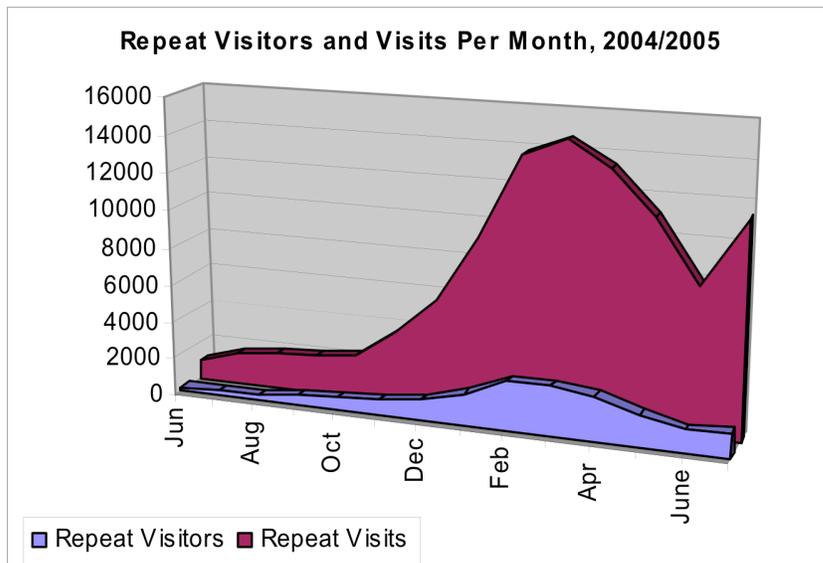


This marketing campaign cost roughly 2% of our project budget.

Web site traffic increased significantly as the outreach activities began and was followed by the expected dip at the end of the school year⁸:



By the end of a 6-month period, a large volume of visits was generated by a growing number of faithful users.



⁸ The following graphs were from the 2005 Q1 – Q2 Usage Statistics for the Women Working Collection and are available through the Internet Archive’s Wayback Machine: http://web.archive.org/web/*hh_/ocp.hul.harvard.edu/WW-Usage-Stats-2005-Q1-Q2.pdf

Visits to the site continued to grow consistently, and by the end of 2005, approximately 13% of the site's visitors were from outside the United States.⁹ Through the outreach campaign we were also able to gather a core group of users to conduct focus groups, online surveys, and interviews. All of this data contributed to subsequent redesigns of the project sites.

'But That's Not Our Job'

The program staff was confronted by significant resistance from the library administration. The most overwhelming comment seemed to be "but that's not our job...we don't do 'marketing'." I would argue that it is precisely the job of the library, though. And without taking responsible measures to let people know that these resources exist, they aren't really accessible.

Less than 10 months into the outreach campaign the program bowed to pressured to reduce costs and to increase the amount of materials digitized. The campaign ended prematurely, and the next collection, *Immigration to the United States* (<http://ocp.hul.harvard.edu/immigration/>) never benefitted from the same attention. The visits to this site were consistently less than half of that of *Women Working* and inlinks to the site remain at approximately one quarter.

There could be any number of reasons for this discrepancy (scope, topic, etc.), but I strongly believe that lack of 'marketing' contributes significantly to an overall lack of accessibility. I also believe that marketing (or outreach as many prefer to call it) is one of the core functions of a library.

Marketing the Library

In 1931, SR Ranganathan wrote:

It is no wonder that, when the library has been extending its scope, changing its outlook and altering its very character and functions, there should not be adequate understanding among the public as to what was been going on.¹⁰

Ranganathan was a mathematician and librarian in India in the first half of the twentieth century. He is most well known for creating the *Five Laws of Library Science* and his work in incredibly prescient.

Ranganathan was adamant—as his first law states—that 'books are for use.' And in support of this law he called for librarians to make their collections known and available,

⁹ Ibid.

¹⁰ S.R. Ranganathan, *The Five Laws of Library Science*, Madras: Madras Library Association, 1931. p. 315.

and to help readers find relevant materials within them, even and especially when patrons “do not know enough about available resources to know what to request.”¹¹

“The majority of readers do not know their requirements, and their interests take a definite shape only after seeing and handling a well-arranged collection of books,” he wrote. He charged librarians to bring related books together—to arrange books in ways that would support the development of the reader’s interests and questions.”¹² He had advice on the best location for a library, “a library which is keen about its books being fully used will plant itself in the midst of its *clientele*,”¹³ the proper height of library shelves, “no rack should be higher than what can be reached by a person of average height, while standing on the bare floor”¹⁴ and the comfort of library furniture. He believed that libraries should stay open longer “In no country, where the concept, ‘BOOKS ARE FOR USE’, has taken root in the Public Mind, will any library be allowed to close till the majority of humanity go to bed and cannot use it”¹⁵ and championed what he called the “open access” movement—open book stacks versus closed ones. His thoughts on librarianship encompassed psychology, sociology, and philosophy:

“In a word, the librarian should be ‘friend, philosopher and guide’ to every one who comes to use the library. It is such sympathetic personal service and ‘such hospitality that makes a library big, not its size’”¹⁶

I speak here about libraries and librarianship because it the core function of academic libraries to support scholarly communities. And when those scholarly communities move onto the web, it is still the function of an academic library to support that them. Libraries are not places for simple information retrieval. They are dynamic spaces for discovery, learning, knowing, and creation. Libraries are essentially there to facilitate the creation of knowledge. They do this by connecting people with information and providing the tools for them to organize, evaluate, and transform it. They are about structuring relationships—between researchers, and between publishers, researchers, information, and knowledge.

¹¹ See Michele V. Cloonan and John G. Dove, “Ranganathan Online: Do Digital Libraries Violate the Third Law?” *Library Journal*, 1 April 2005, <<http://www.libraryjournal.com/article.CA512179.html>> (cited 25 July 2005), para. 4.

¹² Ranganathan cited in Cloonan and Dove, para 5.

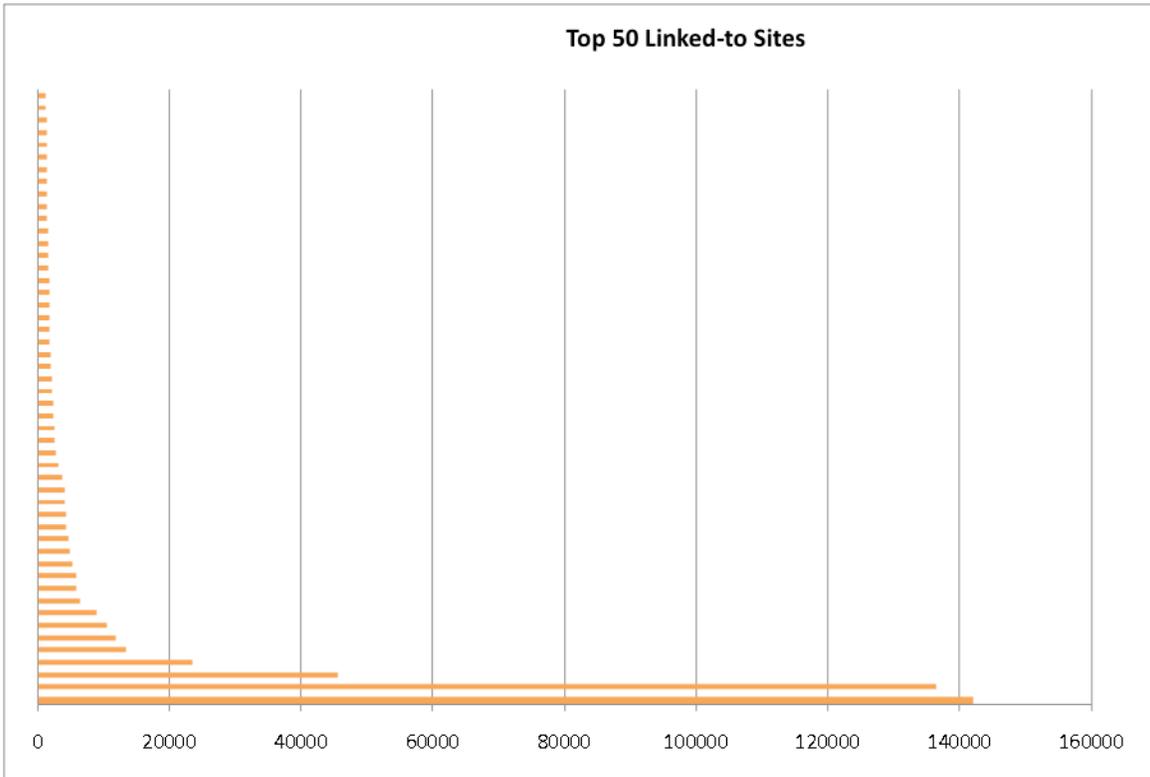
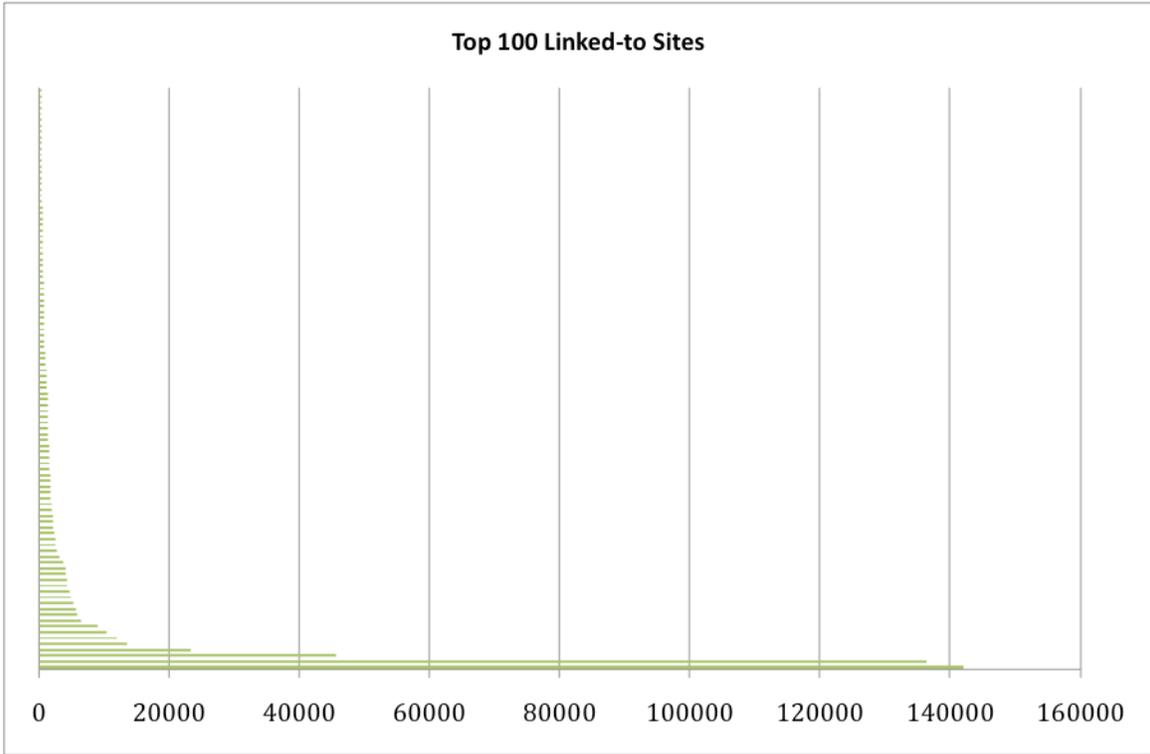
¹³ Ranganathan, p. 10-11.

¹⁴ *Ibid*, p.25

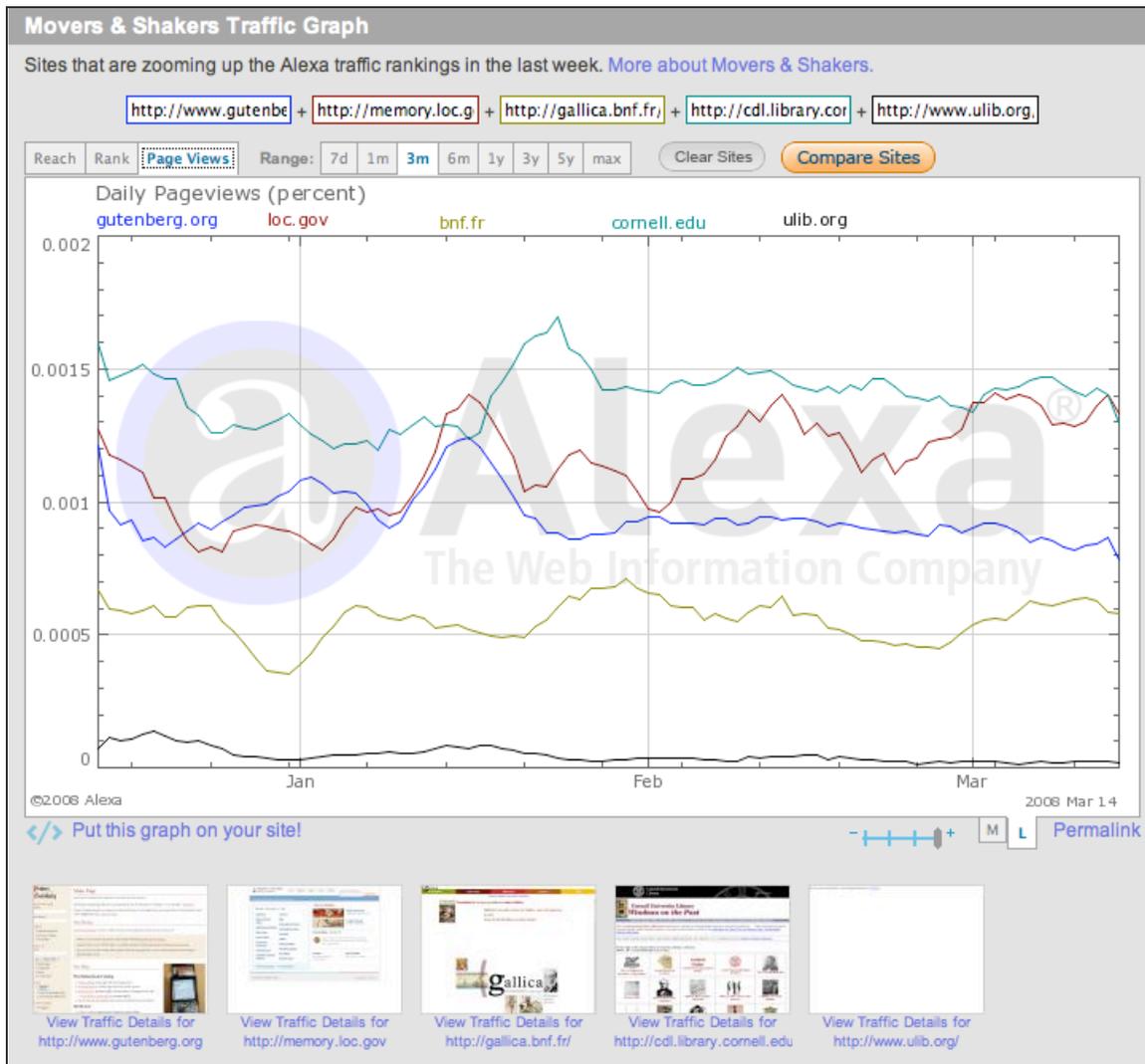
¹⁵ Ranganathan, p. 21

¹⁶ *Ibid*.

Appendix A: Histograms showing the number of inlinks to the top 100 and top 50 most linked-to sites



Appendix B: Alexa the Web Information Company: Graph of Page Views for the Five Most Linked to Sites (3 month period)



1. <http://www.gutenberg.org>
2. <http://memory.loc.gov>
3. <http://gallica.bnf.fr/>
4. <http://cdl.library.cornell.edu/moa/index.html>
5. <http://www.ulib.org/>