**Investigation of the Health and Medicine Blogosphere: Recommendations for Expanding the NLM Health and Medicine Blogs Collection**

# Diana Almader-Douglas, NLM Associate Fellow

July 23, 2013

## Project Sponsors

Christie Moffatt, Digital Manuscripts Program, History of Medicine Division

Kristina Womack, Selection and Acquisition, Technical Services Division

## Table of Contents

Acknowledgments3

Abstract4

Background5

Methodology6

Results13

Discussion and Reflection15

Recommendations17

References18

Appendix I: Original Project Proposal23

Appendix II: Proposed Best Practice Guidelines for Selecting Health and Medicine Blogs25

Appendix III: Health and Medicine Blog Selection Checklist27

Appendix IV: Review of Blog Awards28

Appendix V: Pre-set Proposal Questions29

Appendix VI: Health and Medicine Blog Review ListSeparately Attached

## Acknowledgements

I owe a great deal of gratitude to Christie Moffatt and Kristina Womack, my project sponsors. Christie and Kristina not only submitted a well-organized and clear proposal, they provided a great deal of support, encouragement, and guidance through the course of the project.

A very special thank you to my preceptor, Dianne Babski, for guiding me through the entire year, even during her transition to her position as Deputy Associate Director for Library Operations. My year could not have been completed without her consistent presence and support.

I would like to thank Kathel Dunn, Associate Fellowship Coordinator for her leadership and guidance throughout the year.

I extend gratitude especially to Dr. Lindberg, Betsy Humphreys, Joyce Backus and Dianne Babski for this amazing opportunity to be part of the National Library of Medicine.

Last, I thank my family for believing in me and for everything they did to make sure this was a positive experience for me.

# Abstract

## Objective

The purpose of this project was to expand NLM’s pilot Health and Medicine Blogs collection from 12 to 100 blogs, to more fully represent the medical blogosphere. This process included evaluating the current state of the medical blogosphere and making recommendations for additions to this unique collection of born-digital content. This project also highlights the value of Web content archives.

## Methods

During the project, I:

* conducted a literature review
* researched the blogosphere through an environmental scan
	+ answered pre-formulated questions
	+ investigated how other institutions are approaching Web content collecting
	+ researched the usefulness of health and medicine blog awards in identifying valuable content
* developed a set of best practice guidelines for selecting health and medicine blogs for capture and created a checklist to aid the selection process
* selected and reviewed blogs identified through blog search engines, blog rolls and from ‘word of mouth’ suggestions
* recommended blogs for the NLM Health and Medicine Blogs collection, available at <http://archive-it.org/collections/2722>

## Results

125 blogs were reviewed. Based on the NLM Collection Development Manual, NLM Web Collecting FAQs, the NLM Preservation Policy, and the proposed Best Practice Guidelines (see Appendix II); I recommended 100 blogs to be added to the existing Health and Medicine Blogs collection. 25 blogs were not recommended for addition to the collection.

## Conclusion

NLM is collecting Web content to fulfill its core mission. Through this collection, NLM is capturing health and medicine blogs in a variety of perspectives believed to have long-term value for future scholars.

# Background

NLM’s mission is to acquire, preserve and disseminate biomedical literature. (Collection Development Manual of the National Library of Medicine, 2004). Based on its responsibility to acquire the biomedical literature in any format deemed appropriate to the fulfillment of its mission, NLM decided to expand its capacity to collect Web content.

**Figure 1: Timeline of Web Content Collecting and Health and Medicine Blog Collection at the NLM**

In 2009, a Web Collecting and Archiving Working Group began investigating Web archiving as a method of collection development. The group looked into how other institutions are collecting digital content, and assessed the technical capacity of Web collecting and staffing needs. On March 26, 2010, the group released the ‘Test Physician/Patient Blogs’ Report, in which it established criteria for Web content collection development and presented key findings about challenges of collecting Web content. The group determined the collection scope, addressed challenges of capturing password-protected content, permissions, and changes in blog hosts/URLs. The group evaluated whether links to outside resources should be included in the archive or not. The group also made decisions about crawl frequency and employed methods of quality review.

In January 2011, Julie Adamo completed an Associate Fellow Project titled “Identify Top Bloggers in the Health and Wellness Domains and Recommend Outreach Opportunities for MedlinePlus” and recommended an in-depth review of blog ranking sources (p.6).

In March 2012, the group released key recommendations for Web archiving in a Web Collecting and Archiving Working Group report.

In April 2012, the National Library of Medicine released its pilot collection of Health and Medicine blogs on the Internet Archive’s Archive-It site featuring twelve blogs by physicians, nurses, administrators and patients.

Following the release of the collection, the National Library of Medicine announced the launching of its “Web Content Collecting Initiative to Preserve Health-Related Blogs and Other Media” (U.S. National Library of Medicine, 2012).

A new charter for the Web Collecting and Archiving Group was established in May 2013 to determine workflows and procedures for Web collecting activities at NLM.

# Methodology

I conducted a literature review and researched the health and medicine blogosphere. I also conducted an environmental scan to answer pre-formulated questions presented in the project proposal, investigate how other institutions are approaching Web content collecting, and to research the usefulness of health and medicine blog awards in identifying valuable content.

Based on the findings, I developed a set of best practice guidelines for selecting health and medicine blogs for capture and created a checklist to aid the selection process. Applying these tools, I reviewed and recommended blogs to the Health and Medicine Blogs collection.

## Literature Review

Blogging is an effective and valuable communication tool. In a study investigating communication in an Intensive Care Unit, Curry (2012) found that blogging fosters two-way communication through posts, comments and replies and is a “form of rapid, real-time communication” (p. 328). In her study, Curry identified the advantages of “real-time updates in which important information can be distributed to many readers at the same time” (p. 330). Another study by Chou, Hunt, Beckford, Moser and Hesse (2009) states “the blogosphere presents a tremendous opportunity for health communication” (p.7).

These researchers predict that blogs will be increasingly used for health communication and say that “in the future, social media promise to be a way to reach the target population regardless of socioeconomic and health-related characteristics” and that “social media outlets may represent an excellent opportunity to reach traditionally underserved members of the population” (Chou WY, Hunt YM, Beckford EF, Moser R, Hesse BW, 2009, p. 9). Drawing upon Chou et al., Health Information National Trends Survey (2011) results “suggest social media presents opportunities for communicating with the public about health topics and may hold potential for narrowing the existing health disparity gaps”. Even though Miller and Pole (2010) recognize that blogs may play a role in health communication, they propose that more data about health blogs are needed (p. 1518).

Other findings reveal that for patients, communicating experiences through blogging can be therapeutic and beneficial to authors as well as readers. In a 2012 study, it was found that blogging assisted chronic pain patients “in sharing, responding” and understanding their illnesses and the challenges associated with their conditions (Ressler PK, Bradshaw YS, Gualtieri L, Chui KK, 2012, p. 11). Similar findings were noted by McBride (2011), who stated that blogging creates “opportunities to give and receive emotional support, share experiences and ask questions” (p. 20). Miller and Pole (2010) also found that many health blogs may help patients and caregivers cope with their conditions (p. 1516).

For some patients, the writing itself is beneficial. “Writing provided a vehicle for the bloggers to reflect on their experiences in written form and gain understanding and sense of mastery over their problems” (Marcus MA, Westra HA, Eastwood JD, Barnes KL, 2012, p. 12).

For patients and caregivers, blogs provide emotional support. Patients and caregivers turn to blogs for “interactive support networks” (Miller and Pole, 2010, p.1514). Watson (2012) agrees, “the blogosphere is a resource for patients” (p. 216). According to Gruzd, Blac, Le and Amos (2012), “patients who turn to the blogosphere may be looking more for emotional support and exchange of personal stories than for information about recent research” (p. 41). In some cases, patients turn to blogs because of the anonymity and privacy within these online networks and communities. “According to the researchers, because of the anonymity of the Internet, participants were able to reveal things that they would not discuss in a face-to-facet research setting” (McBride, 2011, p. 20). A study by Kim and Chung (2007) found that “blogs are used more frequently to share emotional support and personal stories than medical knowledge” (p. 449), highlighting the value of emotional support that blogs can bring to patients and caregivers.

Researchers can also benefit from blogs. Watson (2012) reports that blogs “may help increase evidence based practices within institutions via online journal clubs and other strategies” (p. 217).

Blogs have the potential to gain notice, media attention and publicity (McKenzie and Ozler, 2011) and thus have the potential to influence policy formation. Blogs allow people to create virtual venues in which groups of people can become involved, get together, support a campaign and demand change. According to Kovic (2008), “medical blogs are frequently picked up by mainstream media; thus, blogs are an important vehicle to influence medical and health policy” (p.1). Blogs “can actively engage the community in policy-making and provide insight into our decisions” (Rockey, 2013, p. 298).

Blogs can influence healthcare providers, patients, caregivers and researchers. Through blogs, healthcare providers have the opportunity to share important health and medical information with large audiences efficiently. Because of the comment and reply feature of blogs, healthcare providers are able to have dialogue with patients and readers. (Curry, 2012, p. 328-330).

## Environmental Scan

The environmental scan for this project consisted of searching blogs, reports, and series on the Internet for trends, opportunities, or issues affecting blogs or the process of blogging. The environmental scan also included scanning health and medicine specific blogs and blogging practice. From the environmental scan, I gained knowledge of the wide-ranging scope of the blogosphere, the demographics and types of health and medicine bloggers, and blog readership. I learned about the impact of blog awards, and about approaches to identifying valuable content. I also researched Web collecting activities by other institutions.

### The Blogosphere

To gain insight into the current state of the medical blogosphere, it is important to consider the general state of the blogosphere. Blogs, which made their debut in 1994, (Thomson, n.d.) are still highly represented on the Internet and are effective for obtaining and relaying information. According to Bullas (2012),

#### Table 1: Facts about blogs

| **Blog Facts** |
| --- |
| 329 million people view blogs monthly |
| There are 42 million blogs in the US |
| There are over 500,000 daily posts in blogs |
| There are 400,000 daily comments made on posts |

In a Digital Influence Report (2013), Technorati Media revealed that blogs are still powerful sources of information. Swartz (2013) also stated, “blogs are still one of the most influential mediums, ranking high with consumers for trust, popularity, and influence” and are a trustworthy source of Internet information (p. 34).

**Table 2: Demographics of bloggers and health and medicine bloggers**

| **Blogger Demographics** According to Technorati’s State of the Blogosphere (2011) | **Health and Medicine Blogger Demographics According to** Kovic et.al (2008) and Miller and Pole (2010) |
| --- | --- |
| 79% have college degrees | 60% are males |
| Majority are between the ages of 25-44 | 58% are between the ages 30-49 |
| One third are over age 44  | 71-79% of health and medicine bloggers have graduate degrees |

The frequency of how often individuals read health and medicine blogs is unknown, according to Buis and Carpenter (2009), but in the general blogosphere, readership data is available. A 2009 Pew Internet & American Life Project study reported that teens comprise the highest group of blog readers and readership declines with age. In another Pew Internet & American Life Project study, Jones and Fox (2009) report that only 31% of people 73 years of age go online, and 15% of those who go online read blogs (p. 5).

**Table 3: Percentage of blog readership by age**

| **Age**  | **Percentage of blog readers** |
| --- | --- |
| Teens | 49% |
| 18-32 | 43% |
| 33-44 | 34% |
| 45-54 | 27% |
| 55-63 (baby boomers) | 25% |
| 64-72 | 23% |

In yet another Pew Internet & American Life Project study, Lenhart and Fox (2006) reported that 57,000,000 (39%) American adults are reading blogs.

**Figure 2: Percentage of adults reading blogs**

Miller EA, Pole A (2010) analyzed the content and characteristics of health blogs and health bloggers to provide a more thorough understanding of the health and medicine blogosphere and found that about half were written from a professional perspective, one-third from a patient perspective, and a few from the perspectives of caregivers.

**Figure 3: Categories of health and medicine bloggers**

### Web Content Collecting By Other Institutions

Collecting and preserving Web content since 2000, the Library of Congress tested and modeled a variety of capturing mechanisms and tools to develop a strong Web capture infrastructure. Library of Congress has a Web archiving team of curators, or ‘Recommending Officers’, who are responsible for selecting Web content for collections. The process of Web content selection at the Library of Congress is guided by a resource called “Collection Proposal and Specification” and by the Library of Congress selection criteria. Scope is then outlined and submitted to management to determine if resources are available and then are subject to approval processes by Library Management. (Library of Congress Collections Policy Statements Supplementary Guidelines, 2008, p. 1-3). Library of Congress’ efforts in the selection of Web content reflects commitment to Web archiving, and can be a model for other Web-content collecting institutions.

National Archives and Records Administration (NARA) is preserving federal agency Web sites. NARA captured Federal Agency public Web sites in 2004, and captured Congressional Web sites in 2006, 2008 and 2010 and 2012. Although NARA is only crawling Web sites once for one-time snapshots, it uses Heritrix for Web crawling NARA archives Web content that is of national historical value, that protects citizen’s rights or pertains to policy (Conrad, 2013). NARA also had to decide where Web sites end. If the linked information is related, then links are re-written and archived. If the information is out of scope, the links are not archived, and a 404 message appears. NARA provides a disclaimer that notifies the users that the page is an archived Web page (Conrad, 2013). NARA only archives Web content that has integrity, is reliable, authentic, and usable.

This year in Canada, Library and Archives Canada (LAC), a federal institution archiving Canada’s heritage, partnered with Canadiana, an organization of public and research libraries, to digitize the Canadian archival records (Velarde, 2013). The LAC used Heritrix for Web crawling. This is amid controversy over the risk of losing LAC’s digital archives because of budget cuts and leadership changes (CBC News, 2012).

In the UK, the National Archives (NA) is preserving government information published on the Web (National Archives, n.d.). The NA contracted with Internet Archive and harvested 50 Web sites from 1997-2003, and in 2005, contracted with the Internet Memory Foundation to support its Web Continuity Initiative (National Archives, n.d.). NA (n.d.) acknowledges, “Web archiving necessitates continual research into both capture (retrieving and storing data) and accessing (presenting the archived data in the archived site)”.

The National Library of Korea uses “online archiving and searching internet sources” (OASIS) to acquire online resources. To support its selection process, the Library of Korea has an acquisition team, and receives recommendations from members of the public or subject experts. The Web sites are then selected, catalogued, and evaluated for worth and if they are accepted, they are referred to preservation and to the resolution of copyright issues (Conference of Directors of National Libraries in Asia and Oceania, 2009).

### Awards

I reviewed the usefulness of health and medicine blog awards in identifying valuable content, and to find out if awards have influence in the blogosphere. This portion of the project stemmed from former Associate Fellow Julie Adamo’s (2011) suggestion for “a more in depth investigation of blog review and ranking sources” (p. 6). During an environmental scan of health and medicine blog awards, I reviewed nine blog awards. I determined blog awards are not reliable quality indicators for valuable content or ‘good’ blogs (See Appendix IV). With the exception of Technorati Media’s Health and Medicine Top 100 Living/Health Blogs, which rates blogs based on a complex algorithm of rank, and authority (“linking behavior, categorization, and other data” (Technorati.com, n.d.), there was no standard set of criteria for issuing awards to health and medicine blogs. Some awards were based on voting systems in which the same IP address could vote many times in a given time frame, while others were ‘judged’ subjectively by individuals or groups. Some awards were simply suggestions or lists by guest bloggers.

#### Identifying Valuable Content

Upon determining that blog awards were not indicators of quality, I examined the practices and policies within NLM, HHS, and other institutions, which helped identify valuable content. The NLM Wiki contains two key documents that pertain to blogging. The NLM Blogging Policy addresses author transparency, maintenance of the blog, and advises on commenting and replying.

A document in the wiki titled ‘More Information on Blogs at NLM’ provides helpful information about time estimates for posting, commenting and maintenance of blogs. Responding to readers via comments and replies is also addressed.

The United States Department of Health & Human Services (HHS) Center for New Media (HHS Blog Guidance, 2010) provides best practices for blogging and communicates an expectation for quality blogs within the federal government.

The Library of Congress, which is collecting and preserving science blogs, developed criteria for adding science blogs to its collection. One criterion is that the blog must represent “original thought” (Harbster, 2013). Genuine blogs reflect original thought and are not just a collection of links in a blog.

Even health and medicine blogs with fewer updates or small audiences may be highly relevant and useful for the audiences reading them. Smaller health and medicine blogs may be as equally valuable for NLM to collect as more well-known blogs in terms of quality and potential impact on readers and should be candidates for capture. Identifying smaller blogs may present a challenge if they are located in the Deep Web. Deep Web content is “publicly accessible information available via the World Wide Web, but not retrievable using search engines that rely on crawlers or spiders…” (Reitz, 2004).Because less well-known blogs might not be featured in blog search engines, they will probably be best identified through word-of-mouth recommendations, or from blog rolls of other blogs.

# Best Practices Guidelines and Health and Medicine Blog Selection ChecklistI developed best practices guidelines based on criteria from the Web Collecting and Archiving Working Group, the HHS Media suggested blog practices, as well as on the findings in the literature review conducted during this project (See Appendix II). Applying the Guidelines for Selecting Health and Medicine Blogs, I created a selection checklist to aid in the internal process of selecting blogs for the collection.

## Selection Strategy

The goal was to broaden the collection by recommending health and medicine blogs to more fully represent the health and medicine blogosphere.

Guided by the NLM Collection Development Manual, NLM Web Collecting FAQs, NLM Preservation Policy, and the proposed best practice guidelines (see Appendix II); I selected blogs for consideration and added them to an Excel spreadsheet (See Appendix IV).

Blogs were identified utilizing three primary methods.

**Table 4: Methods utilized for identifying blogs to review**

| Methods utilized for identifying blogs to review |
| --- |
| Blog Search EnginesTechnorati Living/Health blogs <http://technorati.com/blogs/directory/living/health/>Best of the Web Blog Directory <http://blogs.botw.org/>Blog Search Engine <http://www.blogsearchengine.org/>EatonWeb Blog Directory <http://portal.eatonweb.com/> |
| Staff suggestions and recommendations |
| Searches of blogrolls |

## Suggested crawl frequency of the health and medicine blogs collection

Crawl frequency is how frequently the browsing and copying of resources using a crawler takes place (International Organization for Standardization (ISO) Working Group, p. 3). For this project, the recommendation for crawl frequency of recommended blogs was based on two primary factors: the options available for crawl frequency in Archive-It, and the frequency of blog posts in each blog.

# Review

In total, 125 blogs were reviewed. I labeled Fields C through I in an Excel spreadsheet (See Appendix V). The fields correspond to signposts for quality, such as transparency of authorship, whether the blog allows comments, whether there are regularly occurring posts. Based on the NLM Collection Development Manual, NLM Web Collecting FAQs, NLM Preservation Policy, and the proposed best practice guidelines (see Appendix II); I recommended 100 blogs to be added to the existing Health and Medicine Blogs collection.

# Results

Thirty-eight of the 100 recommendations were patient blogs. In addition to considering the resources mentioned above, I favored patient blogs that illustrated interesting, unconventional, or fresh perspectives of their experiences with disease, prognosis, and recovery.

Twenty-seven blogs by physicians or other healthcare providers were recommended. I also sought unique and varied perspectives of health and medicine. I favored a balance of professional and personal expression of thoughts and perspectives, to reflect the wide-ranging human nature of healthcare professionals.

In the case of the ten recommended Caregiving blogs, I searched for diverse types of caregivers such as parents of children with special needs, caregivers for patients with dementia, caregivers of aging parents, and caregivers of war veterans. I also tried to locate blogs illustrating adaptive coping skills such as humor and exercise, to capture diverse examples of the human condition.

When reviewing blogs by organizations, I recommended 12 blogs that reflected diverse topics that could be of benefit to wide audiences.

I reviewed five blogs written by students and recommended four. Recommended blogs highlighted education, health, illness, and healthcare from the student perspective. Student blogs were not highly represented in the blogosphere.

I reviewed 17 blogs that did not quite fall into an identified category, or could be of benefit general audiences. Blogs without designated target audiences were less likely to be recommended because the purpose of these blogs tended to be unclear.

Table 5: Blog Selection Recommendations

| Author Type | Recommended For Archive | Not Recommended For Archive |
| --- | --- | --- |
| Patients  | 38 | 9 |
| Physicians/Healthcare Providers | 27 | 3 |
| Caregivers | 10 | 2 |
| Organizations | 12 | 2 |
| Students | 4 | 1 |
| General/Other | 9 | 8 |
| **SUBTOTAL** | **100** | **25** |

Twenty-five blogs were not recommended to be added to the collection because they did not meet the majority of the proposed best practice guidelines for selecting health and medicine blogs (See Attachment IV). Eleven blogs were not recommended because they were out of scope for health and medicine. Four were excluded from recommendation because of technical issues, such as broken links, 404 error messages, non-maintenance, and other technical problems. Five were excluded due to lack of author transparency, and another five were excluded because the blogs contained too many advertisements or were selling and promoting commercial products.

Table 6: Reasons 25 blogs were not recommended

# Discussion and Reflection

Blogs related to health and medicine are being archived and preserved for a different reason than they were created. According to Technorati’s State of the Blogosphere 2011, 70% of individuals blog to share expertise and experience with others. Others blog to influence the ways people think and to express themselves creatively (Kovic 2008).

In contrast, blogs are being archived and preserved for unforeseen uses in a period of time yet to occur (Association of Research Libraries, 2012). It is a challenge to predict what blog content will be considered valuable, so it is important to represent a variety of perspectives for scholars, researchers and historians of the future. By archiving blogs, the NLM is capturing snapshots of human experiences and personal perspectives of health and medicine. Archiving blogs “reflects the human condition” (Harbster, 2013).

According to the NLM Web collecting and Archiving Working Group,

Blogs on topics in medicine have significant research value for historians, medical sociologists, and medical anthropologists of the twenty-first century as they provide insight into the health care environment, practice and attitudes of a given time or place, and document medical innovations and medications. Medical blogs also reflect and may influence health behaviors and viewpoints, and capture dissenting opinions and new trends not found in other publications (Web Collecting and Archiving Working Group, 2010).

Web archiving addresses the volatility and fragility of Web content, in which “valuable and interesting materials often surface and then can quickly disappear” (NLM Web Collecting FAQs, 2012).

Selection of URLs to review for recommendation was a time consuming and labor intensive process. The average amount spent reviewing each blog was about 15-20 minutes. Reviews consisted of checking links , viewed videos, searched blogrolls, scanned archives, read posts, read comments and replies and sought author transparency.

On a large scale, there is a need to be a better way to create efficiencies for selecting content, or to provide more staffing resources. Library of Congress uses an in-house curatorial tool called DigiBoard, for example, is a tool which allows nominated URLs to be stored, viewed and searched (Grotke and Jones, 2010, p. 18) during the selection process.

The selection process for this project was executed by one person, and was therefore subject to some subjectivity. In the future, it is appropriate to assess whether it is suitable for one person to make selection decisions.

Selecting and preserving content from the World Wide Web requires consideration of permissions and copyright issues. I referred to the 2012 Report, which addressed whether or not permission from bloggers or blog owners would be needed to archive and preserve. I also consulted with the project sponsors about the matter. It was decided that capturing Web content for preservation might be considered fair use based on the Association of Research Libraries’ Code of Best Practices in Fair Use for Academic and Research Libraries (2012) Principle #8; “Collecting Material Posted on the World Wide Web and Making It Available” which states “it is fair use to create topically based collections of Websites and other material from the Internet and to make them available for scholarly use” (p. 27). In addition, and in accordance with the ARL Code of Best Practices in Fair Use, I suggested a takedown option and procedure for blog owners in order to strengthen a case for fair use.

Early in the project, I considered selecting Spanish-language blogs to the Health and Medicine Blogs collection. During the blog review phase, I explored whether to include blogs in the Spanish language to the pilot collection. I ran a search in Google Español using search term “blogs de salud” (health blogs). Although the search retrieved qualifying blog for review, I found that categorization may be an issue; other languages were represented in the results; for instance, Diaro de um Bipolar <http://bloggerbipolar.blogspot.com/> is a Portuguese language blog and ‘Dempeus per la salut Pública’ <http://dempeusperlasalut.wordpress.com/>, which was categorized as a Spanish language health blog, was a political blog in the Catalan language. While there is an opportunity to collect blogs in other languages, this will require additional planning and decision-making. Whereas the review of an English blog may take15-20 minutes, reviews of blogs in other languages may require even more time, and may require further review by language specialists to make sure they meet the proposed Best Practice Guidelines. Given the time constraints for this project, I did not make an effort to include Spanish-language blogs in the review, but recognizes that there is potential value to collecting in multiple languages in the future.

Even though cost was not a focus of this particular project, when developing Web content collections, the impact on labor, staffing, time and software is most likely going to be an important consideration.

# Recommendations

The NLM should continue its pioneering quest for the best methods of Web collecting and blog preservation. “Such collections represent a unique contribution to knowledge” and “in the absence of such collections, important information is likely to be lost to scholarship” (ARL Code of Best Practices in Fair Use for Academic and Research Libraries, 2012, p. 26).

In the future, NLM may consider

* enhancing the description of the NLM Health and Medicine Blogs collection Archive-It page
* in addition to the NLM Archive-It Web page, provide access to the NLM Health and Medicine Blogs collection via an NLM site
* developing an in-house curatorial tool to manage selection and seed management
* migrating the collection from Archive-It to its own Digital Collections Repository
* providing metadata for the collection
* promoting the Health and Medicine Blogs collection
* expanding the breadth of born-digital Web collections

# References

Adamo J. National Library of Medicine. Identify top bloggers in the health and wellness domains and recommend outreach opportunities for MedlinePlus. Bethesda (MD): 2011 Jan [cited 2013 Jun 6].

Association of Research Libraries. Center for Social Media, School of Communication, American University and the Program on Information Justice and Intellectual Property, Washington College of Law American University Code of best practices in fair use for academic and research libraries [Internet]. Washington (DC): 2012 Jan [cited 2013 Mar 25]. 24 p. Available from: <http://www.arl.org/bm~doc/code-of-best-practices-fair-use.pdf>

Bullas J. Blogging statistics, facts and figures in 2012- Infographic [Internet]. [Cited 2013 Aug 5]. Available from: <http://www.jeffbullas.com/2012/08/02/blogging-statistics-facts-and-figures-in-2012-infographic/>

Buis LR, Carpenter S. Health and medical blog content and its relationships with blogger credentials and blog host. Health Communication. 2009 Dec 10;24(8):703-10.

CBC News. Federal libraries, archives shutting down [Internet]. 2012 May 2 [cited 2013 Jun 19]. Accessed from: <http://www.cbc.ca/news/canada/ottawa/story/2012/05/02/ottawa-libraries-archives-closing-budget-cuts.html>

Chou WY, Hunt YM, Beckford EF, Moser R, Hesse BW. Social media use in the United States: implications for health communication. J Med Internet Res. 2009 Nov 27;(11)4:e48.

Conference of Directors of National Libraries in Asia and Oceania. CDNLAO newsletter [Internet]. 2009, Nov; cited 2013 Jun 17]. Available from: <http://www.ndl.go.jp/en/cdnlao/newsletter/066/662.html>

Curry K. Increasing communication in the intensive care unit: is blogging the answer? Crit Care Nurs Q. 2012 Oct-Dec;35(4):328-34.

Dig-it8. Importance of the Blogroll [Internet]. [Cited 2013 Jun 7]. Available from: <http://www.digit-8.com/blogging-tips/importance-of-the-blogroll/>.

Farr T. Using a blog to improve communication. Nursing Management. 2011 Jun; 52-3.

Ghosh M. Importance of transparency in blogging [Internet].Weblogtoolscollection.com.2007 Jan 14[cited 2013 May 17]. Available from: <http://weblogtoolscollection.com/archives/2007/01/14/importance-of-transparency-in-blogging/>.

Grotke A, Jones G. DigiBoard: a tool to streamline complex web archiving activities at the Library of Congress. 10th International Web Archiving Workshop; 2010 Sep 22-23; Vienna, Austria.

Gruzd A, Black FA, Le TN, Amos K. Investigating biomedical research literature in the blogosphere: a case study of diabetes and glycated hemoglobin (HbA1c). J Med Libr Assoc. 2012 Jan;100(1):34-42.

Gunelius S. What are blog comments? The importance of blog comments to bloggers [Internet]. [cited 2013 Jun 13]. Available from: <http://weblogs.about.com/od/partsofablog/qt/BlogComments.htm>

Harbster J. Saving science blogs [Internet]. Washington (DC): Library of Congress; 2013 Jan 25 [cited 2013 Apr 15]. Available from: <http://blogs.loc.gov/inside_adams/2013/01/saving-science-blogs/>.

Heritrix Public Wiki [Internet]. [Updated 2012 Jun 12; cited 2013 Jun 17]. Available from: [https://webarchive.jira.com/wiki/display/Heritrix/Users+of+Heritrix](https://webarchive.jira.com/wiki/display/Heritrix/Users%2Bof%2BHeritrix)

International Organization for Standardization (ISO) Working Group. Information and documentation-statistics and quality indicators for web archiving. 2012 Sep 28. 54 p. ISO/DTR 14873.

Jones S, Fox S. Generations online in 2009 [Internet]. Washington (DC): Pew Internet & American Life Project; 2009 Jan 28 [cited 2013 Apr 24]. Available from: <http://www.pewinternet.org/Reports/2009/Generations-Online-in-2009.aspx>

Kim S, Chung DS. Characteristics of cancer blog users. J Med Libr Assoc. 2007 Oct;95(4): 445-50.

Kovic I, Lulic I, Brumini G. Examining the medical blogosphere: an online survey of medical bloggers. J Med Internet Res. 2008 Jul-Sep;10(3):e28.

Ledford J, Teixeira J, Tyler ME. Google analytics. 3rd ed. Indianapolis (IN): Wiley: 2010. 4004 p.

Lenhart A, Fox S. Bloggers: a portrait of the internet’s new storytellers [Internet]. Washington (DC): Pew Internet & American Life Project; 2006 Jul 19 [cited 2013 Apr 24]. Available from: <http://www.pewinternet.org/~/media//Files/Reports/2006/PIP%20Bloggers%20Report%20July%2019%202006.pdf.pdf>

Library and Archives Canada. Government of Canada web archive [Internet]. [Updated 2007 Nov 15; cited 2013 Jun 17]. Available from: <http://www.collectionscanada.gc.ca/webarchives/technical-details/index-e.html>

Library of Congress Collections Policy Statements Supplementary Guidelines, Humanities and Social Sciences Division. Web Archiving. [Updated 2008 Nov; cited 2013 Jun 14). Available from: <http://www.loc.gov/acq/devpol/webarchive.pdf>

Marcus MA, Westra HA, Eastwood JD, Barnes KL; Mobilizing Minds Research Group. J Med Internet Res. 2012 Jan 30;14(1):e17.

McBride D. Kaiser Permanente Oakland Medical Center, Oakland (CA). Cancer survivors find blogging improves quality of life. ONS Connect. 2011 Apr;26(4):20.

Miller EA, Pole A. Diagnosis blog: checking up on health blogs in the blogosphere. Am J Public Health. 2010 Aug;100(8):1514-19.

National Cancer Institute. Health Information National Trends Survey (HINTS). [Internet]. Bethesda (MD): U.S. National Institutes of Health. [2011 Aug; cited 2013 Jun 12]. Available from: <http://hints.cancer.gov/docs/HINTS_Brief_19.pdf>

National Archives. Information on web archiving [Internet]. [Cited 2013 Jun 17]. Available from: <http://www.nationalarchives.gov.uk/webarchive/information.htm>

National Archives. UK government web archive [Internet]. [Cited 2013 Jun 17]. Available from: <http://www.nationalarchives.gov.uk/webarchive/>

Nelson M. Blogging basics 101: What do you want to know? What is a permalink and why do I need it? [Internet]. [2008 Nov 24; cited 2013 Jun 13]. Available from: <http://www.bloggingbasics101.com/2008/11/what-is-a-permalink/>

NLM Associate Fellows (National Library of Medicine, Bethesda, MD). Presentation on Advanced Research by: Mark Conrad (Archives Specialist, Center for Advanced Systems and Technologies (NCAST), National Archives and Records Administration, College Park, MD). 2013 Mar 20.

Reitz, JM. Online dictionary for library and information science [Internet]. Westport (CT): Libraries Unlimited. 2004 [cited 2013 Jul 24]. Available from: <http://www.abc-clio.com/ODLIS/odlis_A.aspx>

Ressler PK, Bradshaw YS, Gualtieri L, Chui KK. Communicating the experience of chronic pain and illness through blogging. J Med Internet Res. 2012 Oct 23;14(5):e143.

Rockey S. Transparency: two years of blogging the NIH. Nature. 2013 Jan 17;493(7432):298-9.

Shore R, Halsey J, Shah K, Crigger BJ, Douglas SP. Report of the AMA council on ethical and judicial affairs: professionalism in the use of social media. J Clin Ethics. 2011;22(2):165-72.

State of the blogosphere 2011 [Internet]. San Francisco (CA): Technorati Media. 2011 Nov 4 [cited 2013 Mar 13]. [about 10 screens]. Available from: <http://technorati.com/social-media/article/state-of-the-blogosphere-2011-introduction>

Swartz J. Technorati Media’s 2013 digital influence report [Internet]. San Francisco (CA): Technorati Media. [2013 Feb 5; cited 2013 Mar 13]. Available from: <http://technoratimedia.com/wp-content/uploads/2013/02/tm2013DIR.pdf>

Technorati.com. Technorati authority FAQ [Internet]. San Francisco (CA). [cited 2013 Jun 13]. Available from: <http://technorati.com/what-is-technorati-authority/>.

Thompson C. The early years [Internet]. New York News & Features. [cited 2013 Jul 22]. Available from <http://nymag.com/news/media/15971/>

U.S. Department of Health and Human Services. HHS Blog Guidance Page [Internet]. Washington (DC): U.S. Department of Health and Human Services; [cited 2013 Apr 16]; [about 4 screens]. Available from: <http://www.hhs.gov/web/socialmedia/getting_started/blog_guidance.html>

U.S. National Archives and Records Administration. Federal web harvests [Internet]. [Cited 2013 Jun 17]. Available from: <http://www.webharvest.gov/>

U.S. National Library of Medicine. Blogging Policy for the National Library of Medicine. [Internet]. Bethesda (MD): U.S. National Institutes of Health; [updated 2010 Nov 30; cited 2013 Apr 30]; [1screen]. Available from: [https://wiki.nlm.nih.gov/confluence/display/NLMBLOGS/Blogging+Policy+for+the+National+Library+of+Medicine](https://wiki.nlm.nih.gov/confluence/display/NLMBLOGS/Blogging%2BPolicy%2Bfor%2Bthe%2BNational%2BLibrary%2Bof%2BMedicine)

U.S. National Library of Medicine. Collection Development Manual [Internet]. Bethesda (MD): U.S. National Institutes of Health; [cited 2012 Apr 15] Available from: <http://www.nlm.nih.gov/tsd/acquisitions/cdm/policy.html>

U.S. National Library of Medicine. National library of medicine preservation policy [Internet]. Bethesda (MD): U.S. National Institutes of Health; [updated 2012 Jul 03; cited 2013 Apr 15]; [1 screen]. Available from: <http://www.nlm.nih.gov/psd/pcm/nlmpres.html>

U.S. National Library of Medicine. NLM launches web content collecting initiative to preserve health-related blogs and other media [Internet]. Bethesda (MD): U.S. National Institutes of Health; [updated 2012 Nov 14; cited 2013 Apr 16]; [about 2 screens]. Available from: <http://www.nlm.nih.gov/news/nlm_web_content_collection.html>

U.S. National Library of Medicine. NLM web collecting FAQs [Internet]. Bethesda (MD): U.S. National Institutes of Health; [updated 2012 Nov 14; cited 2013 Apr 15]; [about 2 screens]. Available from: <http://www.nlm.nih.gov/webcollecting/>.

U.S. National Library of Medicine. Web Collecting and Archiving Working Group [Intranet]. Archive-it collection Test Physician/Patient Blogs Report. 2010 Mar 26. Bethesda (MD): U.S. National Institutes of Health; [cited 2013 Apr 15]. Available from: [https://wiki.nlm.nih.gov/confluence/display/DP/Web+Collecting+and+Archiving+Working+Group](https://wiki.nlm.nih.gov/confluence/display/DP/Web%2BCollecting%2Band%2BArchiving%2BWorking%2BGroup)

U.S. National Library of Medicine. Web Collecting and Archiving Working Group [Intranet]. Summary of Key Recommendations. 2012 Mar 30. Bethesda (MD): U.S. National Institutes of Health; [cited 2013 Apr 15]. Available from:

[https://wiki.nlm.nih.gov/confluence/display/DP/Web+Collecting+and+Archiving+Working+Group](https://wiki.nlm.nih.gov/confluence/display/DP/Web%2BCollecting%2Band%2BArchiving%2BWorking%2BGroup)

Velarde D. No paywall, no privatization [Internet]. 2013 Jun 12; [cited 2013 Jun 17]. Available from: <http://www.canadiana.ca/en/lac-project-faq>

Watson J. The rise of blogs in nursing practice. Clin J Oncol Nurs. 2012 Apr;16(2): 215-17.

# Appendix I

**Project Title**: Evaluation of the current state of the medical blogosphere and recommendations for expanding NLM’s Health and Medicine blog collection

**Submitted By**: Christie Moffatt, Digital Manuscripts Program, History of Medicine Division, Kristina Womack, Selection and Acquisition, Technical Services Division

**Brief Description**:

In April 2012 NLM released a pilot collection of Health and Medicine blogs on the Internet Archive’s Archive-It site at <http://archive-it.org/collections/2722>. The collection features 12 blogs, including the blogs of physicians, nurses, patients, and administrators. The NLM Web Collecting and Archiving Working Group would like to expand this collection to feature 100 blogs, to more fully represent the medical blogosphere. In order to do this, there are several questions that should be researched:

* What is the current state of the medical blogosphere? What do we know from the medical and library literature about who is reading medical blogs?
* Who is blogging about health and medicine? The current collection includes a sampling of doctors, nurses, administrators, and patients. Whose perspectives are missing?
* Who are the top bloggers in health and medicine? Some blogs have been selected for awards, but we need to know more about the criteria used for award. What are some approaches to identifying valuable content?
* What about the importance of health and medicine blogs with fewer updates and smaller audiences? What are these blogs like? Should these voices be captured as well? If they are to be collected, how do we best identify and select these?
* Where possible, assess the impact of medical blogs on health care providers, patients, research, policy formation, treatment and support.

Researching these questions and the preparation of a summary report on findings will shape the Associates work during weeks 1-4 of the proposed project.

During weeks 4-8, the Associate will review and recommend for capture a selection of blogs across the variety of roles and perspectives in health and medicine as identified in the first portion of this project. Ideally, the collection will expand to include approximately 100 blogs.

This project builds on recommendations by Julie Adamo in her 2010/2011 NLM Associate project to “Identify Top Bloggers in the Health and Wellness Domains and Recommend Outreach Opportunities for MedlinePlus.” In her report, Ms. Adamo wrote: “The Associate and project leaders were surprised by both the extent of anonymity and the largely unorganized nature of the blogging world. As we gained a solid understanding of these limitations, we were able to clarify criteria of relevant blogs and move forward (page 5).” She noted that in hindsight, a review of the medical blogosphere would have been valuable in efforts to identify top bloggers, and recommended that along with other surveying methods, “future projects could look more in depth at blog review and ranking sources (page 6).”

It is hoped that this associate project to evaluate the current state of the medical blogosphere and provide recommendations for expanding NLM’s Health and Medicine blog collection will serve as a model for understanding and defining the scope of future NLM Web collecting efforts and developing specific strategies for the identification and selection of content.

**Total Duration:** 8 weeks (4 weeks on the evaluation of medical blogosphere, 4 weeks on the identification of health and medicine blogs to capture)

**Primary Learning Objectives for Associate**:

The Associate will gain a deeper understanding of the nature and extent of the medical blogosphere, and a working knowledge of the challenges and complexities of identifying Web content for collection and long-term preservation.

**Expected Project Experiences**:

Project scope definition

Examination of an unfamiliar technical area

Development and presentation of recommendations

**Expected Outputs/Products**:

A report that identifies the major groups of individuals who are blogging in health and medicine (doctors, nurses, patients, medical students. . . who else?); developing a set of criteria for collecting blogs including consideration of the impact of these blogs; and lastly, an assessment of the long-term value of preserving health and medicine blogs and a recommendation about what role NLM should play in this preservation activity.

Recommendations for the addition of approximately 100 new blogs across identified blogging groups to NLM’s Health and Medicine Blogs collection.

**Suggested Methodologies**:

Literature Review, including literature about the medical blogosphere and the impact of medical blogs, NLM collecting guidelines, and Web collecting guidelines at other libraries

Environmental Scan of medical blogosphere to understand and summarize the range of content, perspectives, and roles

Review of Web sites to inform recommendations for NLM collecting

**Benefits to NLM:**

Documentation by the Associate of the scope of content to be collected and the processes and challenges of selection will aid in the development of effective workflow procedures to optimize the quality of Web collecting at the National Library of Medicine.

The current Health and Medicine Blogs collection will expand from 12 blogs to 100, and more fully represent the medical blogosphere.

**Project Leader(s):** Christie Moffatt, Digital Manuscripts Program, History of Medicine Division, Kristina Womack, Selection and Acquisition, Technical Services Division

**Other Resource People**: Jennifer Marill, Jeff Reznick, Paul Theerman

NLM Web Collecting and Archiving Working Group including the initial project wiki site at [https://wiki.nlm.nih.gov/confluence/display/DP/Web+Collecting+and+Archiving+Working+Group](https://wiki.nlm.nih.gov/confluence/display/DP/Web%2BCollecting%2Band%2BArchiving%2BWorking%2BGroup)

**Software/server access required**: Internet Access, Microsoft Office

# Appendix II

**Best Practice Guidelines for Selecting Health and Medicine Blogs**

1. At least 50% of the blog posts should be related to medicine or health to maintain the scope of the collection.

2. There should be at least 20 posts to a blog before it is selected; the number of posts gives indication of author commitment to a blog.

3. Blog posts should occur on a regular basis, with no large or unexplained gaps between posts; large or unexplained gaps may indicate lack of commitment to a blog.

4. Authorship should be transparent.

a. There should be a clearly defined person or group of people responsible for the blog.

b. There should be a method for which to contact the blog ‘owner’.

c. The purpose of the blog should clear.

The willingness to share and be contacted about what is shared supports trustworthiness and credibility of online authors. [[1]](#footnote-1)

5. The blog should be genuine and reflect original thoughts. Blogs full of links to other blogs, ads, or spam should not be archived.

6. The information in the blog should be useful and have research value.

7. There should be a Comment section in the blog, as well as a method for replying to comments. Maintaining a Comment section demonstrates engagement to readers allows for online discussions that might not be available elsewhere.[[2]](#footnote-2)

8. There should be a blogroll (list of links in the sidebar of the blog, linking to other blogs) present. A blogroll indicates a network within the blogosphere. Bloggers use blogrolls as even exchanges for linking to one another with the purpose of having more visitors possibly visit a blog.[[3]](#footnote-3)

9. Blogs authored by healthcare providers must protect patient privacy, maintain patient/provider boundaries, and adhere to the patient-provider relationship in online settings.[[4]](#footnote-4)

10. Although varying opinions are desired in the blog and in the comment section, no content should be derogatory or offensive toward others.2

11. An attempt should be made to notify blog owners that their blog is being crawled, archived and preserved.

12. There should be a take-down option and procedure for blog owners who do not want their blogs archived in the NLM Health and Medicine Blogs collection.

# Appendix III

**Health and Medicine Blog Selection Checklist**

| Frequency of blog posts |
| --- |
|  | Twice or more daily |
|  | Daily |
|  | Weekly |
|  | Monthly |
|  | Bi-monthly |
|  | Quarterly |
|  | Semiannually |
|  | Annually |
| Transparency | YES | NO |
| There is a clearly defined author or group of authors responsible for the blog. |  |  |
| There is a method to contact the blog owner. |  |  |
| The purpose of the blog is clear.  |  |  |
| Value to the NLM | YES | NO |
| Blogs show original thought and observations. |  |  |
| Is the information in the blog useful? |  |  |
| Will this blog supplement an existing collection? |  |  |
| Does the blog have potential research value? |  |  |
| Authority  | YES | NO |
| Blogs authored by healthcare providers do not contain personal health information of patients. |  |  |
| If there are comments, comments are free from derogatory or offensive statements toward others. |  |  |
| Author commitment to the blog | YES | NO |
| The blog has at least 20 posts. |  |  |
| At least 50% of the blog posts are related to health or medicine. |  |  |
| Blog does not have large or unexplained gaps between posts. |  |  |
| Network | YES | NO |
| There is a blogroll or list of recommended blogs or resources in the sidebar of the blog.  |  |  |
| Indication of Readership | YES | NO |
| There is a ‘Comments’ and/or ‘Reply’ section in the blog.  |  |  |
| Recommend for archive? | YES  | NO |
| If not recommended for the archive, please provide a reason. |

# Appendix IV

**BLOG AWARDS** **Table: Review of Blog Awards**

| AWARD | CRITERIA |
| --- | --- |
| Best Health Blogs of 2010: <http://www.foxnews.com/health/2010/12/22/best-health-blogs/>  | Undefined method of “search” for award-winning blogs |
| 2010 Medical Weblog Awards on MedGadget sponsored by Epocrates and Lenovo <http://www.medgadget.com/2011/01/the_2010_medical_weblog_awards_finalists_sponsored_by_epocrates_and_lenovo.html> | Nomination and voting process in which one evaluator made a decision “based on a merit and own internal voting results” |
| E drug search Healthcare 100 <http://www.edrugsearch.com/healthcare100/algorithm> | Algorithm: Google Points based on PageRank, number of delicious bookmarks of that blog, tweet count, Facebook likes and shares, # of stumbles, LinkedIn points, thumbsUp Points as well as E drug’s subjective ranking and frequency of updates |
| Wellsphere Top Health Blogger Wellsphere.com award “top HealthBlogger” | Not stated |
| Healthline’s Best Health Blog of 2012 Contest <http://www.healthline.com/health/best-health-blogs-contest> | Contest of popularity; anyone can nominate a blog. Nominations are anonymous; votes are cast on Twitter and Facebook. Users can vote once every 24 hours; blog with most votes wins |
| Forbes Best Medical Blogs <http://www.forbes.com/2003/10/03/cx_mh_1003medblogs.html> | Rated on “presentation, quality, relevance, reputation”  |
| Top 10 Clinical Medicine Blogs (blogs.com) 2010 <http://www.blogs.com/topten/top-10-best-clinical-medicine-blogs/> | Guest provides his/her top 10 list |
| Top Psychology Blogs of 2012: Blog rank <http://www.medgadget.com/2011/01/the_2010_medical_weblog_awards_finalists_sponsored_by_epocrates_and_lenovo.html> | 20 factors: RSS feed membership yahoo incoming links, Google page rank, monthly visitors, pages per visit, social sites popularity.  |
| Technorati Media <http://technorati.com/what-is-technorati-authority> | Linking behavior, categorization, social media activity |

# Appendix V

## Preset Questions in Project Proposal

### What is the current state of the medical blogosphere, and what do we know from the medical and library literature about who is reading medical blogs?

### Who is blogging about health and medicine?

#### Who are the top bloggers in health and medicine?

#### Some blogs have been selected for awards but we need to know more about the criteria used for the award.

#### What are some approaches to identifying valuable content?

#### What about the importance of health and medicine blogs with fewer updates and smaller audiences?

#### Should these voices be captured as well?

#### If they are to be collected, how do we best identify and select these?

#### Assess the impact of medical blogs on:

#### Health care providers

#### Patients

* Research

#### Policy formation

#### Treatment and support

1. Ghosh M. Importance of transparency in blogging [Internet].Weblogtoolscollection.com.2007 Jan 14[cited 2013 May 17]. Available from: <http://weblogtoolscollection.com/archives/2007/01/14/importance-of-transparency-in-blogging/> [↑](#footnote-ref-1)
2. U.S. Department of Health and Human Services. HHS Blog Guidance Page [Internet]. Washington (DC): U.S. Department of Health and Human Services; [cited 2013 Apr 16]; [about 4 screens]. Available from:<http://www.hhs.gov/web/socialmedia/getting_started/blog_guidance.html> [↑](#footnote-ref-2)
3. Dig-it8. Importance of the Blogroll [Internet]. [Cited 2013 Jun 7]. Available from: <http://www.digit-8.com/blogging-tips/importance-of-the-blogroll/> [↑](#footnote-ref-3)
4. Shore, R, Halsey, J, Shah, K, Crigger, BJ, Douglas, SP. Report of the AMA council on ethical and judicial affairs: professionalism in the use of social media. J Clin Ethics. 2011;22(2):165-72. [↑](#footnote-ref-4)